IN SILICO DEVELOPMENT OF BROAD SPECTRUM ANTIBACTERIAL BY TARGETING PEPTIDE DEFORMYLASE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology

In

Biotechnology

By

ASIF KHURSHEED

211BM2005

Under The Supervision of

Dr. BIBHUKALYAN PRASAD NAYAK

Department of Biotechnology & Medical Engineering

National Institute of Technology

Rourkela-769008, Orissa, India

2013

Dr. Bibhukalyan Prasad Nayak
Assistant Professor
Department of Biotechnology & Medical Engineering
National Institute of Technology, Rourkela, Orissa, India

Certificate

This is to certify that thesis entitled "***In silico* Development of Broad Spectrum Antibacterial by Targeting Peptide Deformylase**" by **Asif Khursheed (211 BM 2005)** submitted to the National Institute of Technology, Rourkela for the Degree of Master of Technology is a record of bonafide research work, carried out by him in the Department of Biotechnology and Medical Engineering under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/ Institute for the award of any Degree or Diploma.

Dr. Bibhukalyan Prasad Nayak

Assistant Professor

Department of Biotechnology and Medical Engineering

NIT Rourkela, 2013

Acknowledgement

If words are reflected as symbols of appreciation and token of acknowledgement, then words play the role of thanks to exhibit the deeply embedded feeling of gratitude. I am greatly indebted to, who either through guidance, discussion or providing facilities for the thesis work, have served as a beacon light or crowned my efforts with success. With an overwhelming sense of pride and genuine obligation I express my deep sense of heartfelt gratitude and regards to my guide **Dr. B.P Nayak**, Department of Biotechnology and Medical Engineering for giving me an opportunity to do the project work in an independent arena.

I consider it a privilege to express my gratefulness to **Mr. Raghunath Satpathy**, Department of Biotechnology, MITS, Rayagada for the valuable guidance, and suggestions during the project work.

Further I would like to express my thankfulness to **Mr. Tarun Agarwal**, **Mr. Rashmiranjan Behera** and **Mr. Ekalabya Bissoyi**, for their constant support and advice in my project work. It is an immense pleasure to thank my friends **Arshad Jamil**, **Girish Kumar Sahu, Ajay Kawade**, **Patitapabana Parida, Akshaya Kuma Padhi, Vinod Vishwakarma, Mahaveer Prasad Gupta, Ghalib Nashtar** and all others for their persistent encouragement and day-to-day support.

I sincerely thank all my faculty members for their blessings.

Finally I am grateful to my parents **Mr Khursheed Alam**, **Mrs Naseema Khursheed** and family members for their endurance, love, wishes and support, which helped me in completion of my work. Above all, I thank **Allah** who showered his blessings upon us.

(**Asif Khursheed**)

LIST OF CONTENTS

ABSTRACT

Peptide Deformylase (PDF) is an essential metalloenzyme and removes the formyl group from methionine at the N terminus of nascent polypeptide chains. Since this protein is essential for survival of most of the pathogenic bacteria and functionally equivalent gene is absent in mammalian cells, it provides an attractive target for development of novel antibacterials. In the current project, an *In silico* approach has been used to develop a novel broad spectrum antibacterial based on binding affinity of PDF with derivatives of hydrazines. Virtual screening and Molecular docking based on Lamarckian Genetic Algorithm was carried out to find binding affinity of PDFs from seven different pathogenic bacteria. These molecules were screened from PubChem database on the basis of structural similarity to sulfonylpiperidine. Molecular Dockings were carried out for each ligand and results were clustered together with accepted RMSD of 0.5 A^{o}. Hydrogen bond analysis was done on UCSF Chimera while ADME and toxicity properties were evaluated using Molsoft and PreADMET server. 1-(4-fluorophenyl) sulfonylpiperidine-3-carbohydrazide (CID_4539974) was found to be the best potent inhibitor of PDF, made H-bonds with active metal binding amino residues, successfully passed the Lipinski rule of five, gained good drug-likeness model score, and exhibited negative carcinogenicity and toxicity test. Further, molecular dynamics simulation (MDS) was done to check the stability of protein-ligand complex under using GROMACSv4.05. Negative system energies of protein-ligand complex, negligible deviation of C_{α} from crystal structure and decrease in fluctuations in region containing active metal binding amino acid residue supported the stability of protein-ligand complex. The outcomes concluded that CID_4539974 have all the desirable properties to become a novel broad spectrum antibacterial by targeting bacterial PDF efficiently that needs further validation by wet lab experiments.

Keywords: PDF, Sulfonylpiperidine, AutoDock, Docking, MDS.

List of Figures

List of Tables

Abbreviations

fs	Femtosecond
GA	Genetic Algorithm
LGA	Lamarckian Genetic Algorithm
MD	Molecular Dynamics
MDR	Multi Drug Resistant
MRSA	Methicllin Resistant *S. aureus*
nm	nanomolar
PDB	Protein Data Bank
PDF	Peptide Deformylase
PDI	Peptide Deformylase Inhibitor
ps	picosecond
SA	Simulated Annealing

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1 Introduction

If we see in the past, we find that the life of human being is always endangered by infectious disease. However, in the 21st century prosperous societies live in fantasy that they are free from contagious diseases, moreover, if they any how attacked by pathogenic bacteria, there will be antibiotic to cure the disease. But, if we do not find alternatives to increasing antibiotic resistance, the present day scenario will be false. The discovery of antibiotics is one of the greatest findings of the 20th century. At that time the infectious diseases were the leading cause of global fatality, the introduction of sulfa drugs and penicillin into medical use in 1930s and 1940s respectively led to the treatment of infectious disease and decreased mortality rate [1]. After the success of penicillin, intensive screening of antibacterial agents led to the discovery and successful clinical development of erythromycin, rifamycin, vancomycin, chloramphenicol, cephalosporin, tetracycline, streptomycin between 1940 and 1960 [2]. Despite all advances in the field, problems arise in the proven ability of bacteria to adapt and to develop resistance to multiple classes of existing antibiotics. Infectious diseases are still the second-leading cause of death worldwide and the third-leading cause of death in developed countries and become the global public health fear [3-4]. While a large group of antibacterials to conflict bacterial diseases are presently available, a number of frights have recently been raised on the need to develop new antibacterials. This need has arisen for several reasons; among them is the expansion of antibacterial resistant bacterial infections, rising of infectious diseases not yet recognized, re-emerging of infectious disease experienced previously that have reappeared in more virulent forms and the consequential high and economic impact of infectious disease. The emergence of antibiotic resistance bacteria is a question of not "if," but "when" will the antibiotic resistance occur. To get a sense of the seriousness of this matter, early resistance to multiple drugs was reported among the enteric bacteria, namely, *Escherichia coli*, *Shigella* and *Salmonella*, in the late1950s to early 1960s [5-7]. Recent resistance has been reported in the nosocomial *S. aureus* strains, both methicillin-resistant (MRSA) [8] and multidrug-resistance (MDR) [9]. It now seems that the emergence of antibiotic-resistant bacteria is inevitable to most every new drug. These

problems pose an urgent challenge to find new approach to combat these antibiotic resistant bacteria. To prevent the emergence and dissemination of resistant bacteria; continuing efforts to develop new antibacterial agents are needed i.e. by blocking or circumventing the resistance mechanism or identifying novel drug targets.

1.2 Bacterial Diseases

Bacterial diseases comprise any type of infection or illness caused by bacteria. They are microscopic organism and cannot be seen by naked eye. Millions of bacteria are normally inhabitant the human skin, intestine and genital parts. Most of the bacteria's do not cause disease, however some are even good for health. Disease causing bacteria's are called pathogenic bacteria and they cause infection by reproducing themselves with growth rate greater than that of good bacteria's inside the body and producing toxins that damage the body. Some common bacteria and the types of disease they cause are-

Leptospira interrogans belongs to obligate aerobe eubacteria family, with periplasmic flagella. It is sometimes transmitted to humans through contact with infected animal excretory product, either directly or in water. It invades directly through broken skin and may replicate in the liver and kidneys. Infection is thought as leptospirosis, which might be symptomless, a gentle non-specific sickness, or end in death from liver harm and nephrosis (Weil's syndrome).

Escherichia coli (***E. coli***) belongs to gram -ve, rod-shaped microorganism that's ordinarily found within the lower intestine organ of warm blooded organisms (endotherms). Most E. coli strains are harmless; however some serotypes will cause tract infection, serious food poisoning in humans.

Enterococcus faecium is a Gram-positive, spherical cell that can found in pairs or chains alpha hemolytic or nonhemolytic bacterium in the genus Enterococcus. It can be commensal and can inhabitant within the human intestine; however it may also even be pathogenic, and inflicts diseases like surgical wound infection, nosocomial bacteremia, endocarditis, tract infections and neonatal meningitis.

Staphylococcus aureus is a member of the Firmicutes, and is usually inhabitant the human respiratory tract and skin. Disease-associated strains typically promote infections by manufacturing potent protein toxins. Also they express cell-surface proteins that bind and inactivate antibodies. Although *S. aureus* is not always pathogenic, it is frequent cause of respiratory illness, skin infections and sickness. Moreover, the emergence of antibiotic-resistant forms of pathogenic *S. aureus* (e.g. MRSA) becomes a worldwide downside in clinical medication.

Streptococcus mutans belongs to facultative anaerobes; gram-positive coccus-shaped microorganism ordinarily found within the human oral cavity and may be a vital contributor to dental caries.

Pseudomonas aeruginosa is a common bacterium that may cause illness in animals, as well as in humans. They are gram-negative, asporogenous, rod-shaped, and monoflagellated bacteria that have a fantastic nutritionary versatility. The flexibility permits the organism to infect broken tissues or those with compromised immunity. The symptoms of infections are generalized inflammation and sepsis. If such colonizations occur in vital body organs, like the lungs, the urinary tract and kidneys, the results can be fatal. Because it can flourishes on variety of surfaces, this bacterium is also found on and in medical instrumentation, as well as catheters, inflicting cross-infections in hospitals and clinics.

Vibrio cholerae belongs to family of Gram-negative bacteria and is comma-shaped with one, polar flagellum which helps in movement. There are some strains of *V. cholerae*, which are pathogenic in nature. *Vibrio cholerae* infects the human intestine and enhance mucous production, causes diarrhea and vomiting that lead to dehydration and finally death, if not treated. It is typically transmitted through the excreta of Associate in nursing infected person, usually by approach of unclean water or contaminated food.

The examples discussed above are only few in thousands of pathogenic bacteria causing different types of diseases. Moreover the emergence of drug resistant bacteria raises the essential need for approaching computational method to reduce the time for discovering new drugs.

1.3 Computational Drug Discovery

Computational drug discovery provides good strategy for fast and economizing drug discovery and development method. It minimizes time, efforts and cost of drug discovery and development. Recent advances in bioinformatics and impressive increase in the availability of both biological macromolecule and small molecule information has brought us too considerably nearer to computational drug discovery.

In silico strategies will facilitate in distinguishing drug targets via bioinformatics tools. They will even be accustomed to identify the target proteins and analyze their structures for possible binding sites. *In silico* method help in production of candidate molecules, confirmation for their drug resemblance, docking of these molecules with the target protein, ordering them consistently with their binding energies affinity, and further optimizing the molecules to enhance binding characteristics.

The utilization of complementary experimental and informatics techniques increases the prospect of success in several stages of the discovery procedure, from the identification of novel targets and revelation of their functions to the discovery and development of lead compounds with desired properties. Computational tools supply the benefit of delivering new drug candidates more quickly and at a lower price. Chief roles of computation in drug discovery are; Virtual screening, *in silico* ADME/T prediction & de novo design and advanced strategies for determining protein-ligand binding.

1.4 *In Silico* target identification

The *in silico* approach for novel drug target identification depends on genomic data. Genome information are helpful in the recognition, validation, choice of the potential candidates and screening based on "essentiality" and "selectivity" criteria of the microbial systems [10]. These subtractive genomics approach are used for choosing non homologous protein coding gene, found in pathogen but not in host. These genes are crucial for the existence of pathogenic bacteria and not present in human. This will cut back the chance of cross reactivity and side effects. In this case, Peptide Deformylase has been selected as a novel drug target on the basis of

essentiality and selectivity criteria. This protein is essential for bacterial survival, removes formyl group from nascent peptide and lead to protein maturation. There is no risk of cross reactivity because in human, equivalent functional PDF protein is not yet reported.

The protein executes its function through interaction with other molecules such as ligand, substrate, DNA and other domains of proteins. The 3-D structure of protein provides the mandatory shape and physicochemical texture to facilitate these interactions. Structural data of protein surface regions enables detailed studies of the relationship of protein structure and function. The very first step to obtain a good result should be the identification and visualization of protein cavities for structure-based drug design (SBDD) applications. The activities of protein exist in the cavities where binding of substrate lead to triggering some incident, such as a conformational change or chemical modification. There are various server are present on internet that can be use for finding active site. CASTp, pocket finder, Q site finder, PASS are some available server used for identifying active site in our target PDF.

To identify a reliable and potent drug for protein target protein-ligand autodocking are used. The aim of docking is to predict the binding energy of the protein-ligand complex given the atomic coordinates. Recent enhancements in search algorithms and energy functions, computational docking strategies became a valuable tool to explore the interaction between protein and its inhibitors. The binding energy between the protein and its ligand is calculated by a simplified, often grid-based force field [11].

Active sites of a protein are key factor for the flexible docking. Autodock4.2 is an automated docking tool that predicts how ligand molecules bind to receptor of known 3D structure and it also optionally enables to model binding parameters of ligand with number of distinct conformational clusters and to find all possible minimum binding energy.

1.5 Objectives

The primary objectives of current project are:

1. To identify the different pathogenic bacteria's expressing PDF protein and to perform multiple sequence alignment to find the similarity between PDF proteins.
2. To identify a new class of molecule that has potential to inhibit PDF.
3. To analyze the derivatives (from PubChem) of selected molecule for their potential to be broad spectrum antibacterial drugs.
4. To analyze and dock all chosen molecules for the identification of novel drug candidates.
5. To study molecular, ADME, toxicity properties and MD simulations of novel drug candidates.

1.6 Plan of Work

Plan of Work	3rd Semester		4th Semester	
	Mid Semester	End Semester	Mid Semester	End Semester
Literature review				
Identification a new class of molecule that has potential to inhibit PDF				
Analysis of derivatives of identified molecule and docking with whole set of PDF				
Study of molecular, ADME, toxicity properties and MD simulations of novel drug candidates.				
Manuscript writing				

CHAPTER 2

LITERATURE REVIEW

2. LITERATURE REVIEW

2.1 Function of peptide deformylase

The prokaryotic and eukaryotic organelle (mitochondrial and chloroplast) ribosomal protein synthesis universally initiates with an N-formylmethionine residue[12]. Peptide Deformylase (PDF) is an essential metalloenzyme that removes the formyl group from methionine at the N terminus of nascent polypeptide chains followed by protein maturation. This protein is expressed in all pathogenic bateria, together with Mycoplasma and Chlamydia species. The bacterial PDF is encoded by "def gene". functionally equivalent gene is not found within mammalian cells [13]. This "def gene" is a vital for bacterial growth and survival [14-15]. The catalytic domain of PDF contains three extremely conserved motifs and belongs to the matrix metallo-protease (MMP) family of enzymes. The PDF enzyme requires a single ferrous ion for its activity and catalyzing the reaction

N-formyl-L-methionine + H_2O = formate + methionyl peptide.

In all prokaryotes, a formyl-methionine residue is inherently incorporated at the N-terminus of all nascent polypeptides; however, the vast majority of the N-formyl moiety is removed from the mature proteins. The removal of the N-formyl moiety is catalyzed by Peptide Deformylase (PDF) [16-17]. But, significantly only half of all the polypeptides retain the methionine at their N-terminus [18]. The removal of the N-terminal methionine residue from the nascent polypeptides is catalyzed by the action of methionine aminopeptidase (MAP) [18]. PDF plays, at least, a critical role in mediating the maturation process of the emerging polypeptides, partially due to the requirement of removing the N-formyl group to render the nascent polypeptides for MAP cleavage of the N-terminal methionine residue. Blockade of bacterial Peptide Deformylase produces inhibition to protein synthesis. In fact, the disruption of the PDF encoding gene in *E. coli* proved critical for cell survival [14, 19], making it an attractive drug target.

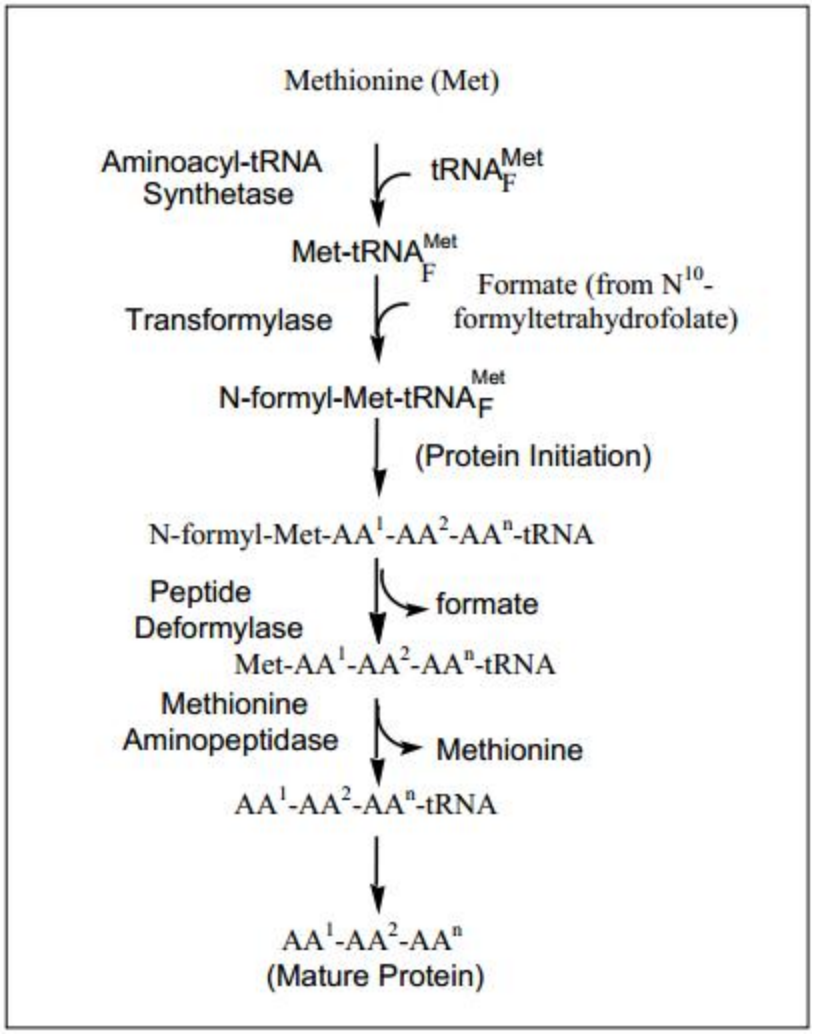


Figure 1: Protein biosynthesis pathway in bacteria, mitochondria, and plastid. Met- methionine; AA- amino acids.

2.2 Structure of Peptide Deformylase

The three-dimensional structure of PDF has been determined for various bacterial species. Moreover, several enzyme–inhibitor complex structures were also resolved [20]. The structure of zinc-containing *E. coli* PDF (Ec-PDF) was at first solved by NMR methods and by X-ray crystallography [21-22]. The 3-D structure of PDF revealed that the metal ion (Fe^{2+}) is tetrahedrally coordinated by Cys90, His132, His136 and a H_2O molecule. Since Fe^{2+} is sensitive to oxygen so for ease of enzymatic study, the ferrous-containing PDF can be replaced by a

divalent metal such as Ni^{2+} [23] or Co^{2+} [24] to make a highly stable and fully active enzyme, as well as that enables spectroscopic studies. A worthy note here is that the enzyme, in both nickel-containing and cobalt-containing PDF, is oxygen insensitive and catalytically as active as the ferrous-containing PDF. From *S. aureus* the PDF have isolated (SaPDF) and its crystal structure was determined in complex with its natural inhibitor actinonin at 1.90 Å resolutions by Hye-Jin Yoon et al. [25]. The three-dimensional structures of the PDF with synthetic inhibitors disclose that the active-site metal ion is pentacoordinated by the metal-binding ligands of the inhibitors [26].

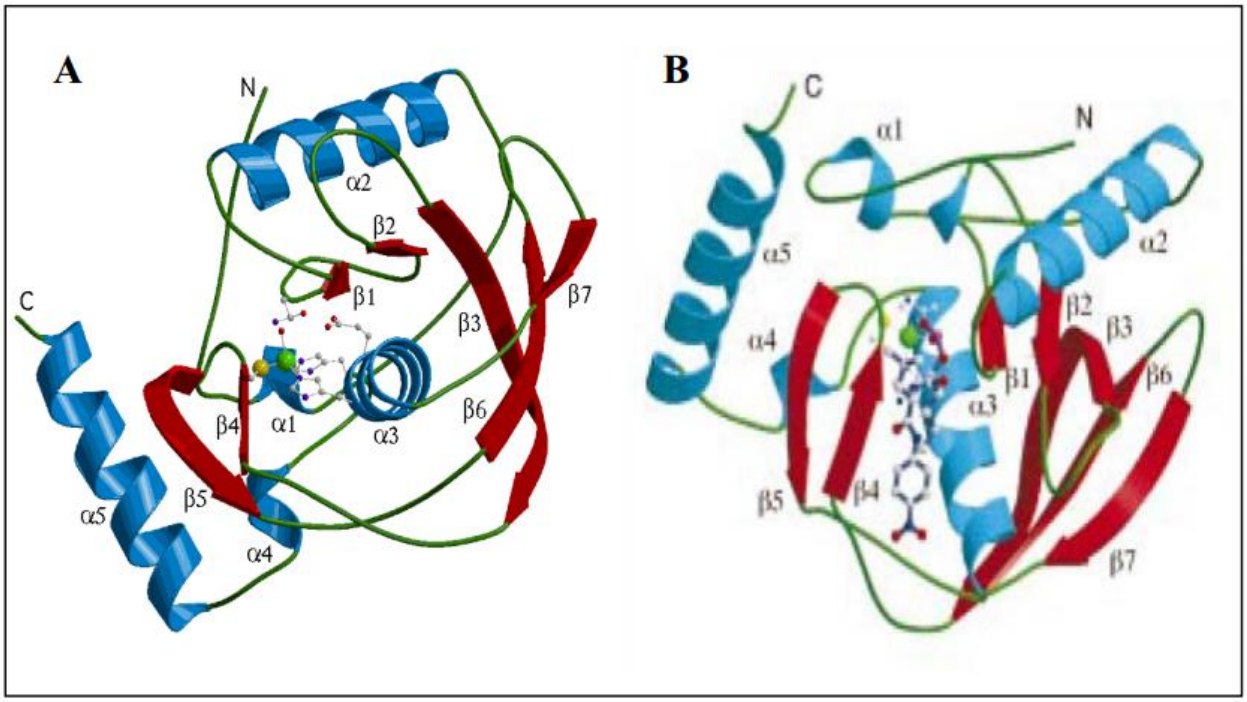


Figure 2: Crystal structures of PDF (A) *E.coli* Zn-PDF structure, (B) *E.coli* Zn-PDF-PCLNA complex

2.3 Peptide Deformylase in Eukaryotes

Protein synthesis in the cytoplasm of eukaryotes doesn't begin with N-formylmethionine, and undoubtedly the need for deformylation in the cytoplasmic process is apparently deficient. However, PDF like homologs are revealed in eukaryotes. In addition to plants a number of plastid containing parasites such as *Plasmodium falciparum* shown to have nuclear encoded gene

homolog [27]. But from sequence alignment, *P.falciparum* PDF shows a 33% identity to the bacterial PDF, alongwith around 60 amino acid extensions at the N-terminus. This demonstrates that, the extra extension is a bipartite N-terminal pre-sequence that consists of signal peptide for entry into the secretary pathway and a plant-like transit peptide for subsequent import into the apicoplast of the Plasmodium. Moreover, PDF-like sequences have been recognized in the genomes of Drosophila Melanogaster, as well as in the partial expressed sequence tags (ESTs) of mouse and human [28]. Like PfPDF, all of these eukaryotic PDF sequences contain an extra ~60 amino acid N-terminal extension. Finally in eukaryotes, these PDF-like gene sequences are not known to be expressed and, if expressed at all, to encode a functional PDF [12]. PDF exists in three forms: PDF1A, PDF1B and PDF2. PDF1B and PDF2 are found in bacteria only and PDF1A is found in human (mPDF) [29]. Therefore, inhibitors selective for bacterial Peptide Deformylase do not show inhibitory activity against mPDF. The bacterial Peptide Deformylase inhibitors (PDIs) are very selective and show activity against similar mammalian (including human) enzymes only at extremely high concentrations which not seems to be achievable in vivo [15]. This has shown by various studies. For instance, the Peptide Deformylase (Ni complex) of *S. pneumoniae*, *Haemophilus influenzae*, *E. coli*, and *S. aureus* was inhibited by BB-81384 with IC50 values of 9, 11, 60, and 300 nM, respectively [8] whereas equivalent IC50 values for the human metalloenzymes collagenase (MMP-1), gelatinase (MMP-2) and angiotensin converting enzyme (ACE) were found to be 10000, 60000, and 5000 nM, respectively [8] (corresponding IC50 refers to the concentration inhibiting enzyme activity by 50 percent). Moreover it was found that in the development of VRC 3852, 20 out of 21 related compounds (95%) showed IC50 values of 100 nM or less for E. coli PDF and 18 out 20 (90%) showed IC50 values of 75 nM or less for *S. pneumoniae* peptide deformylase. But in the case of human, all 21 compounds were very selective for PDF, with IC50 values for tested human metalloenzymes being 200000 nM or greater [9].

2.4 Peptide Deformylase as Novel Drug Target

Since PDF is required for bacterial survival but apparently unnecessary in animal cells, it provides an attractive target for a novel antibacterial strategy. Moreover, deformylation is a conserved feature throughout the eubacterial kingdom. With the numerous three-dimensional

structures of various bacterial PDFs solved [30-34], both with and without enzyme-bound inhibitor complexes, it appeared that the active site of PDF is conserved, composing of sequence motifs, motif 1 [GφGφAAXQ], motif 2 [EGCLS], and motif 3 [HEφDH] (where φis a hydrophobic amino acid and X is any amino acid) [26, 35]. Consequently, PDF appears to be a broad spectrum novel antibiotic drug target.

One notable advance in PDF inhibitor screening came from the random screenings by Versicor and Hoffmann-La Roche [36-37]. The work led to the identification of actinonin, a naturally occurring hydroxamic acid derivative known to have potent antibacterial activity and it is a PDF target. In vitro activity of actinonin against PDF is within Ki value of 0.3 nM and demonstrated bacteriostatic activity against a wide spectrum of Gram-positive and Gram-negative bacteria. Evidence to support the molecular target for antibacterial activity of actinonin came from the *E. coli* genetic construct in which mutant E. coli strain expresses controllable levels of PDF [37]. Thus, the susceptibility to the presence of actinonin is associated with the different expressed levels of PDF within the bacterial cell (at high PDF expression level, the bacterial cell becomes more tolerable to the presence of actinonin) [37]. Despite having strong inhibition for PDF, actinonin was never developed for the treatment of infections due to its poor bioavailability and consequent lack of in vivo efficacy [38].

Peptide deformylase (PDF) inhibitors have long been sought as antimicrobial agents. The discovery of inhibition of PDF by actinonin oriented research efforts toward peptidomimetic, hydroxamic acid-based inhibitors. Hence, the general structure of most current PDF inhibitors contains a strong chelating moiety such as a hydroxamic acid. Owing to a structural lead from actinonin, BB-3497 was developed to target PDF (KI= 9 nM) with moderate antibacterial activity against a range of bacterial pathogens followed by oral administration [39]. N-alkyl urea hydroxamic acids were also shown have antibacterial activity against PDF [9]. Also results showed that hydrazides could give better results as an inhibitory drug against PDF [40].

Moreover the PDF is present in diverse group of eubacteria. This will allow the identification of broad spectrum antibiotics. The aim of this project is to identify more chemical compounds showing anti bacterial properties against PDF and to study their molecular, ADME, and toxicity properties along with pharmacophore modeling and Molecular Dynamics (MD) simulation of protein-ligand complex to analyze the stability under biological conditions.

2.5 Novel drug design strategy

Drug discovery and development is highly complex and, lengthy and process. For a new ligand to reach in the market as a potent drug must develop by the process which is commonly known as developmental chain or "pipeline: and consists of a number of distinct stages [41]. Generally it can be grouped under two stages, early pharmaceutical research and late pharmaceutical R&D. Early pharmaceutical research involves two steps process in which identification and modeling of the biological target within the body (the protein) is the first step, followed by the second step of identification of lead compound (ligand) that shows drug-like properties with respect to this protein. Later, the drug goes through many phases of clinical development in humans. In the clinical stage, the drug is administered first to animals and then to human volunteers to determine:

- The channel of the drug through the body- starting the time when it is administered in the body to the time when it is excreted from the body.
- Drug's effects on the body.
- Its efficiency to cure the disease being treated.
- Adverse side effects of the drug.

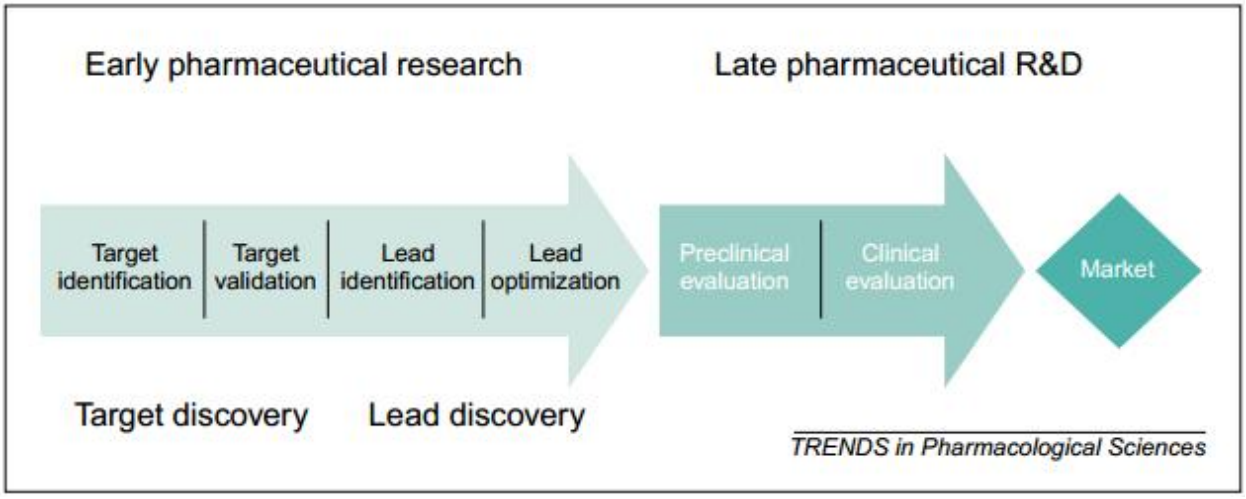


Figure 3: Drug discovery and development process.

However, failure of most candidate molecule can occur as a result of combination of reasons such as poor pharmacodynamics, poor pharmacokinetics, side effects, lack of efficacy and

commercial reasons. Also the cost and time required for the development of new drug from conventional method is very high. From conventional methods it would take approx $1.8 billion and 15 years from the initial stage to the successful marketing of a new drug [42]. Most drugs are discovered by either modifying the structure of known drugs, by screening compound libraries or by developing proteins as therapeutic agents [42]. With the advent of genomics, proteomics, bioinformatics and technologies like, NMR, crystallography, the structures of more and more protein targets are becoming available. So, there is an urge need for novel computational tools for drug designing that can identify and analyze active sites and suggest potential drug molecule that can bind to these sites and not only shorten the R&D time cycle but also reduce the ever increasing cost involved in the drug discovery process. *In silico* drug design or CADD (Computer Aided Drug Design) method fills this research requirement. Moreover, PricewaterhouseCoopers states that "the overall cost of drug development could be reduce by as much as 50% through extensive use of *in silico* technologies in drug discovery procedure" [43]. Also, studies from quantum mechanics, molecular dynamics, molecular dockings, QSAR to ADMET prediction including dissolution studies are performed *in silico*.

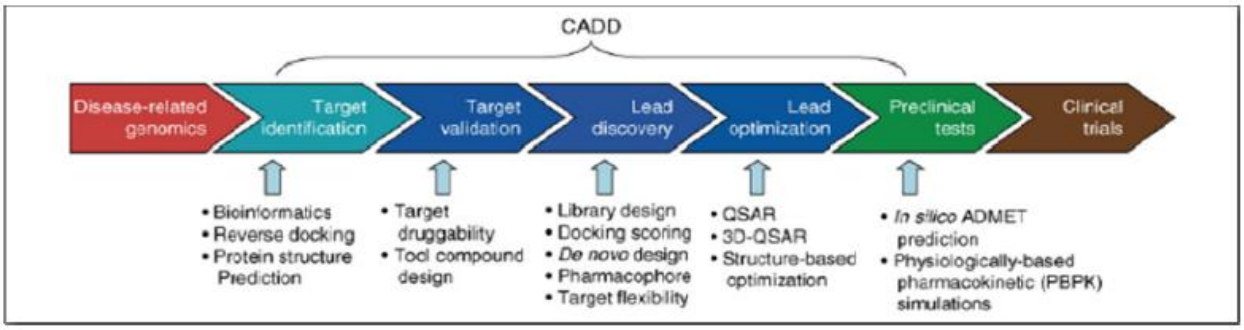


Figure 4: Application of *in silico* drug design (CADD) in various stages of drug development.

In silico drug discovery process comprise of three stages:

(i) First stage involves identification of a therapeutic target and building a heterogeneous small molecule library to be tested against it, followed by the development of a virtual screening protocol initialized by either docking of small molecule from the library or building these structures in the active site by *De novo* design method.

(ii) In second stage, these selected hits are checked for specificity by docking at binding sites of other known drug targets.

(iii) In third stage, these selected hits are subjected to detail *in silico* ADMET profiling studies and those molecules that pass these studies are termed as leads [44]. These lead molecules can be further carried through drug discovery and development cycle.

2.6 Methods of virtual screening in drug design process

The shift from conventional to genomics and proteomics based drug design research has fundamentally transformed key R&D strategies in the pharmaceutical industry addressed to the design of new chemical entities as drug candidates against a variety of biological targets. Therefore, drug discovery has encouraged toward more rational strategies based on our increasing understanding of the fundamental principles of protein-ligand interactions [45]. Virtual screening is generally referred to as the process of scoring and ranking molecules in large chemical libraries according to their state of having affinity for a definite target, [46].

Mainly, there are two methods of virtual screening: ligand-based and target-based. Analysis and comparison of ligand-based virtual screening and target-based virtual screening imply that the knowledge about a target obtained from known bioactive ligands is as important as information of the target structures itself for identifying novel bioactive scaffolds through virtual screening [47-48]. Hence, the absolute selection for a method to use will finally depend on the type and amount of information available without prior having a large impact on performance.

2.6.1 Ligand-based virtual screening

There exist a broad range of ligand-based virtual screening methods. The extent of complexity, and thus the final computational cost of these methods, depends greatly on the type of structural information of ligand being used [49]. Each and every one of these methods rely on the fundamental similarity-property principle according to which similar molecules should exhibit similar properties [50] and therefore chemical similarity calculations form the foundation of ligand-based virtual screening [51]. Thus, all the molecules in a particular database can be scored comparative to the similarity to one or multiple bioactive ligands and then ranked accordingly to reflect decreasing possibility of being active.

2.6.2 Target-based virtual screening

The availability of structural information of the target protein, that being either determined experimentally or derived computationally by means of homology modeling techniques is the core of Target-based virtual screening methods [52-53]. The aim of these methods is to provide, on one side, a good approximation of the expected conformation and orientation of the ligand into the protein cavity (docking) and, on the other side, a reasonable estimation of its binding affinity (scoring). The goal is to select from existing compound libraries or to design compounds with three-dimensional complementarity (i.e. shape, size and physicochemical properties) to the target-binding site. Advancement and new approaches can directly guide the design of virtual combinatorial libraries, which are first tested *in silico* for target complementarity. This will reduce the number of compounds that have to be synthesized and tested in vitro. Target based virtual screening is most popular and gained a reputation in successfully identifying and generating novel bioactive compounds.

2.6.3 Diversity based virtual screening

The aim of diversity based virtual screening is to select a smaller sub-library from a larger compound library in such a way that the full range of chemical diversity is best represented [54]. When the structural information about the target or ligand or both is unavailable, diversity design is the method of choice.

2.7 Docking of protein and drug molecules by AutoDock tool

AutoDock is widely used in the prediction of bimolecular complexes for structure & functional analysis and in molecular design. It combines an empirical free energy force field with a Lamarckian Genetic Algorithm (LGA), providing fast prediction of bound conformations with predicted free energies of association [55].

AutoDock was released in 1990 [56] and consistently from early time it has proven to be an effective tool. It can predict accurately and quickly the bound conformations and binding

energies of ligand with macromolecular targets [55, 57]. The primary algorithm used by AutoDock for conformational searching is the Lamarckian Genetic Algorithm (LGA) [55]. The "Lamarckian" allows individual conformations to search their local conformational space, to find local minima, and ultimately pass this information to later generations. AutoDock4 is also incorporated with simulated annealing (SA) search method and a traditional genetic algorithm (GA) search method. It uses a semi empirical free energy force field to predict binding free energies of small molecules to macromolecular targets [57] and shows a standard error of about 2–3 kcal/mol in prediction of binding free energy in cross-validation studies.

AutoDock Tools (ADT) is inbuilt with object-oriented programming language Python which is made from reusable software components [58-59]. ADT exists in the context of a rich set of tools for molecular modeling, the Python Molecular Viewer (PMV) [59-60]. PMV is a freely distributed Python-based molecular viewer. ADT consist of a set of commands dynamically extending PMV with commands specific for the preparation, launching and analysis of AutoDock calculations. Hence all PMV commands are naturally available in AutoDock tools.

2.8 Molecular dynamics simulation: a method for validation of docking result

Molecular Dynamics (MD) simulation is a theoretical physics technique used for the analysis of molecular system at atomic level and has sound basis in statistical mechanics and classical physics [60-61]. MD simulation is a method for time scale generation of conformations, thermodynamic information, system total energy with kinetic and potential energy of proteins and other biological macromolecules [62-64]. MD simulation provides fine detail concerning the motions of individual particles as a function of time and can be utilized for the quantification of properties of a system at a precision and on a time scale that is otherwise difficult to get to. Therefore, in extending our knowledge of model systems simulation plays as a valuable tool. An important aspect of MD simulation is that theoretical consideration of a system additionally allows one to investigate the specific contributions to a property through "computational alchemy" [65] i.e., modifying the simulation in a way that is nonphysical but nonetheless allows a model's characteristics to be probed. The artificial conversion of the energy functions from that representing one system to that of another during a simulation represents an example of

computational alchemy. This technique is important and use for free-energy calculations [66]. Hence, MD simulation with its range of complementary computational advancement, has proven to be a valuable tool for exploring the basis of protein structure and function [67].

CHAPTER 3

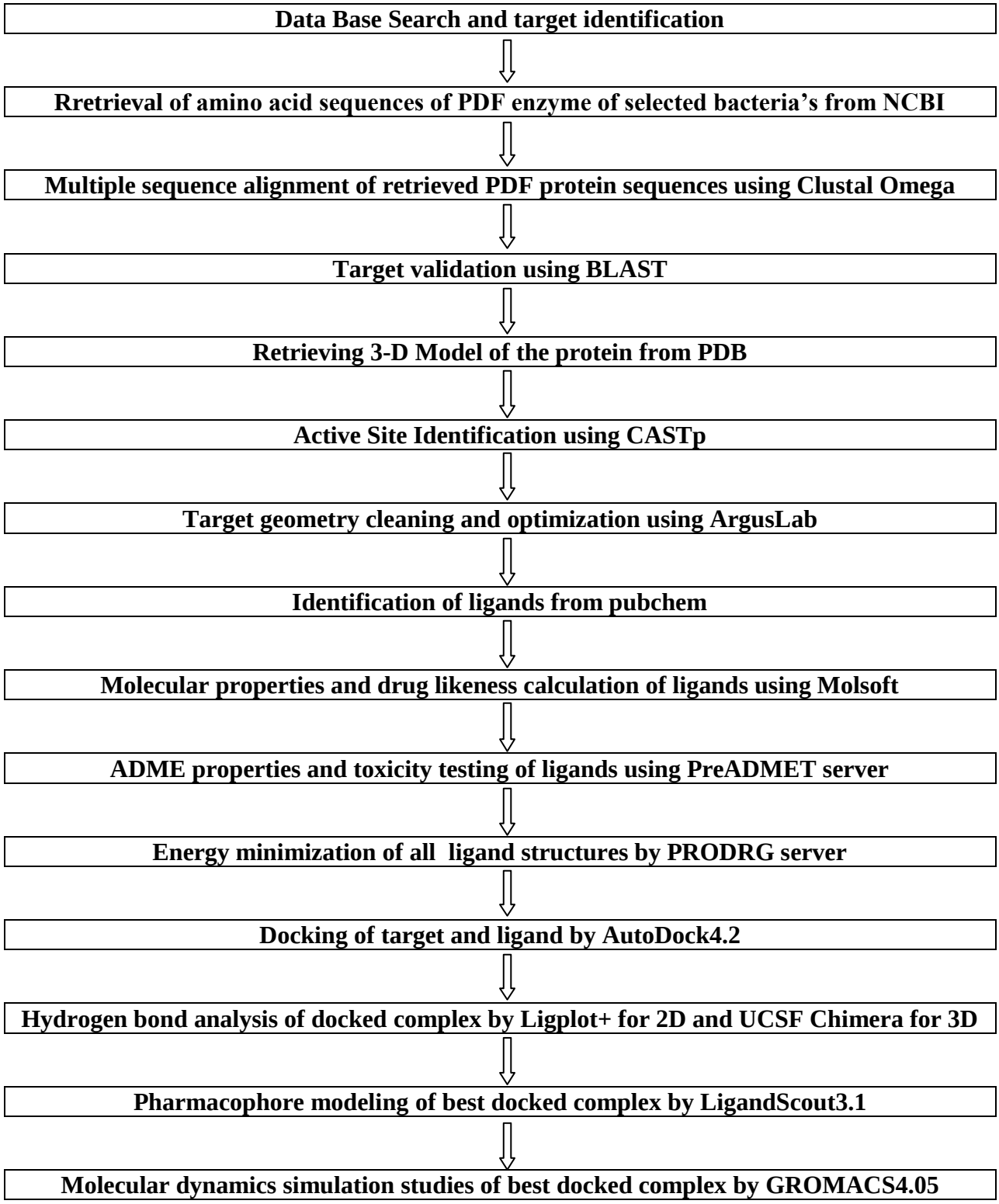
MATERIALS AND METHODS

3. MATERIALS AND METHODS

3.1 Bioinformatics softwares and tools used

1. NCBI
2. Clustal Omega
3. BLAST
4. Protein Data Bank (PDB)
5. CASTp
6. PubChem
7. ArgusLab
8. Molsoft
9. PreADMET
10. PRODRG2
11. AutoDock4.2
12. Open babel 2.3.1
13. LigPlot+
14. UCSF Chimera
15. Rasmol
16. PyMOL
17. LigandScout3.1
18. GROMACS4.05
19. Marvin Space

3.2 Flow chart of protocol followed



3.3 Data Base Search and target identification

After literature survey the database search was carried out to find microorganism expressing PDF protein. The protein data bank is the repository for 3-D structural data of large biological molecules, such as nucleic acids and proteins. These structural data are usually obtained by X-Ray Crystallography and NMR spectroscopy. From the protein data bank 35 different bacteria expressing PDF protein were found. In these we select following threaten pathogenic bacteria for further study-

- *Leptospira interrogans*
- *E.coli*
- *Enterococcus faecium*
- *Pseudomonas aeruginosa*
- *Vibrio cholerae*
- *Staphylococcus aureus*
- *Streptococcus pneumoiae.*

3.4 Retrieval of amino acid sequences of PDF enzyme of selected bacteria's from NCBI

National Centre for Biotechnology Information (NCBI) was established on Nov 4, 1988 as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). It provides an inclusive website that features biological databases and tools for viewing and analyzing the information inherent within the databases. It allows researchers, biologist to access and retrieve biological information from the website. NCBI constructs automatic systems for storing and analyzing information regarding biology, organic chemistry, and genetics; facilitating the employment of such databases and software system by the research and medical community; coordinating efforts to collect biotechnology information each nationwide and internationally; and performing research into advanced strategies of computer-based informatics for analyzing the structure and performance of biologically necessary molecules.

- The URL http://www.ncbi.nlm.nih.gov was open.

- In the search option “Protein” was selected and name of protein & bacteria was written for retrieving amino acid sequence.
- The sequence format was set to FASTA.
- The amino acid sequences of PDF for each bacterium were retrieved and saved.

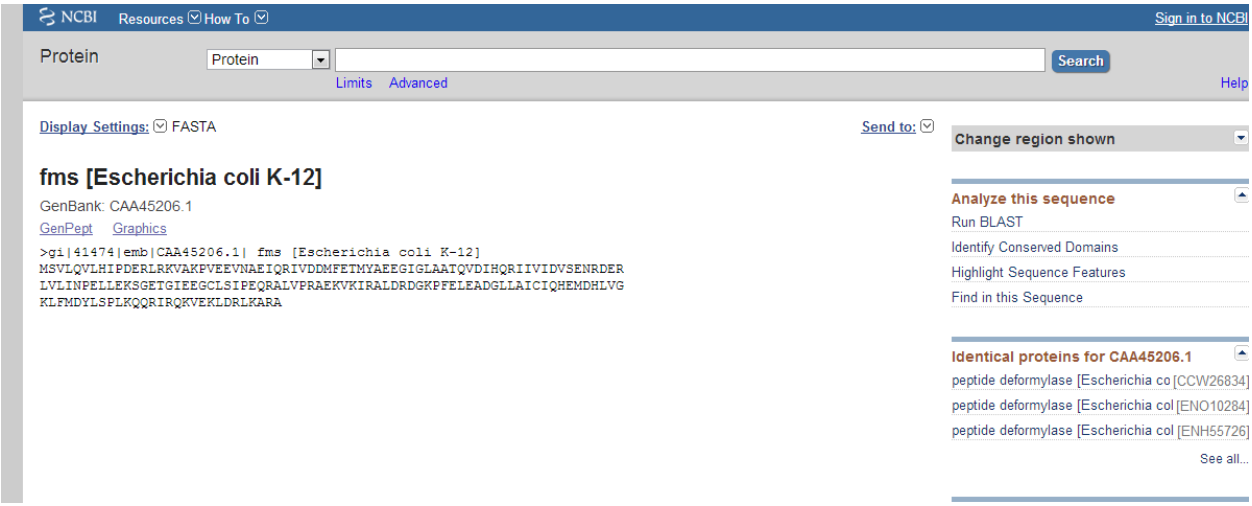


Figure 5: Screenshot of FASTA format of Peptide Deformylase (*E.Coli)* on NCBI web page.

3.5 Multiple sequence alignment of retrieved PDF protein sequences using Clustal Omega

Clustal Omega is a tool developed and maintain by European Bioinformatics Institute (EBI). It produces biologically significant multiple sequence alignments (MSA) and calculates the most effective match for selected sequences and features them up, thus identities, similarities and variations can be identified.

Multiple sequence alignment of PDF protein from the entire selected organism was performed in order to calculate the percentage identity between them. The tool Clustal Omega used for this purpose was available on internet.

- The URL http://www.ebi.ac.uk/Tools/msa/clustalo was opened.
- The FASTA format of all retrieved PDF protein sequences was prepared in notepad.
- In step 1 the set of protein sequences were paste in “enter input sequences” option bar.
- The default parameters were set in step 2.

- The job was submitted in step 3.

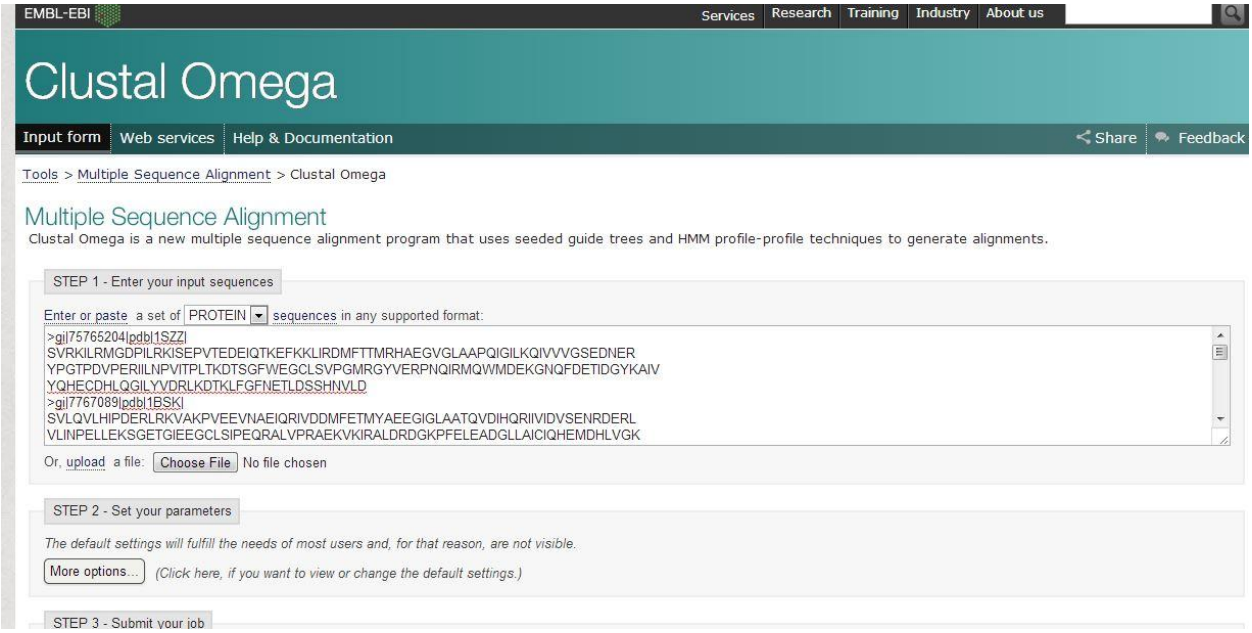


Figure 6: Screenshot of multiple sequence alignment tool Clustal Omega.

3.6 Target validation using BLAST

PDF enzyme is encoded by the def gene, and is present in all disease causing bacteria. Functionally equivalent gene is not present in mammalian cells. Therefore, the target protein is valid.

Moreover for the confirmation of absence of functionally equivalent gene of PDF in human, protein BLAST was run. It stands for Basic Local Alignment Search Tool. It is an algorithm that compares primary biological sequences such as protein and DNA and is provided by NCBI.

- The URL http://www.ncbi.nlm.nih.gov was accessed to open the NCBI homepage.
- BLAST option was clicked.
- For comparing protein sequences pBLAST was selected.
- In the "query sequence" search option FASTA format of bacterial PDF was entered.
- In "choose search set" option database was set to non redundant protein sequences and organism was set to *Homo sapiens* (taxid: 9606).

- Algorithm was set to blastp (protein-protein BLAST).
- Program was run and the result was saved.

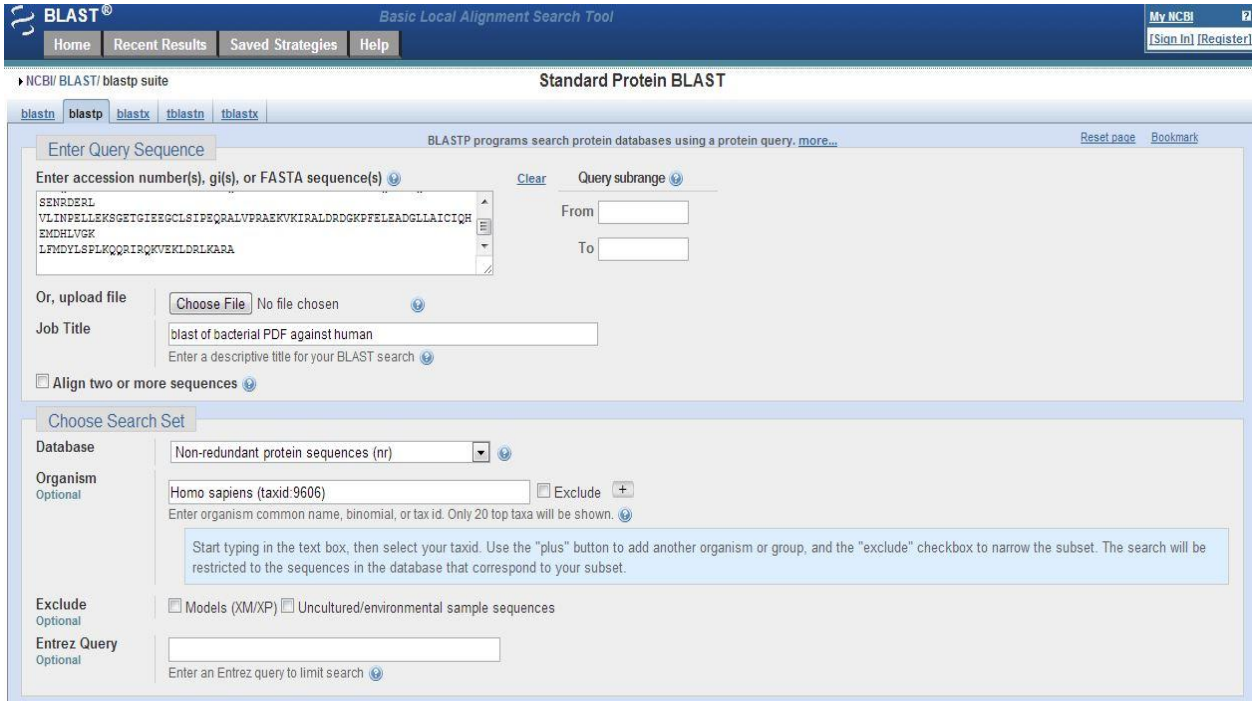


Figure 7: Screenshot of BLAST homepage.

3.7 Retrieving 3-D Model of the protein from PDB:

After validating the target protein its 3-D structure have been downloaded and studied. The 3-D structure of the protein was obtained in pdb format from protein primary database structure PDB.

- The URL http://www.rcsb.org/pdb/home/home.do was opened.
- The four letter pdb code of protein was entered in the search option.
- Files were downloaded in pdb format and saved.

Figure 8: Screenshot of Protein Database (PDB) webpage.

We download the respective pdb file of Peptide Deformylase (PDF) of selected bacterium. The name of microorganisms, their respective PDF PDB Id and download link is listed below-

Table 1: List of microorganism and their respective PDF PDB Id and download link

S.No.	PDB ID	Organism	Protein	Reference
1	1N5N	*Pseudomonas aeruginosa*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=1N5N
2	1BSK	*Esherichia coli*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=1BSK
3	3G6N	*Enterococcus faecium*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=3G6N
4	1SZZ	*Leptospira interrogans*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=1SZZ
5	3L87	*Sreptococcus mutans*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=3L87
6	3QU1	*Vibrio cholerae*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=3QU1
7	1LQW	*Staphylococcus aureus*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=1LQW
8	1XEN	*Esherichia coli*	Peptide Deformylase	http://www.rcsb.org/pdb/explore/explore.do?structureId=1XEN

3.8 Active Site Identification using CASTp

Sites of activity in proteins usually lie in cavities. It is the size and shape of protein cavities that dictates the three-dimensional geometry of ligand so as to fit likes a hand in glove. The binding of a substrate typically serves as a mechanism for chemical modification or conformational change of protein. Binding sites are often targeted by various ligands in attempts to interrupt related molecular processes. Active sites of a protein are a key factor for the flexible docking. Active sites of the Peptide Deformylase (PDF) were predicted by using tool CASTp (computed atlas of surface topography of proteins). CASTp calculates concave surface regions on three-dimensional structures of proteins. For this it takes into account of pockets located on protein surfaces and voids buried in the interior of proteins that are frequently associated with binding events. Also, it calculates the size of mouth openings of individual pockets for better accessibility of binding sites to various ligands and substrates. The result obtained made more precise by pocket Finder and Q site Finder.

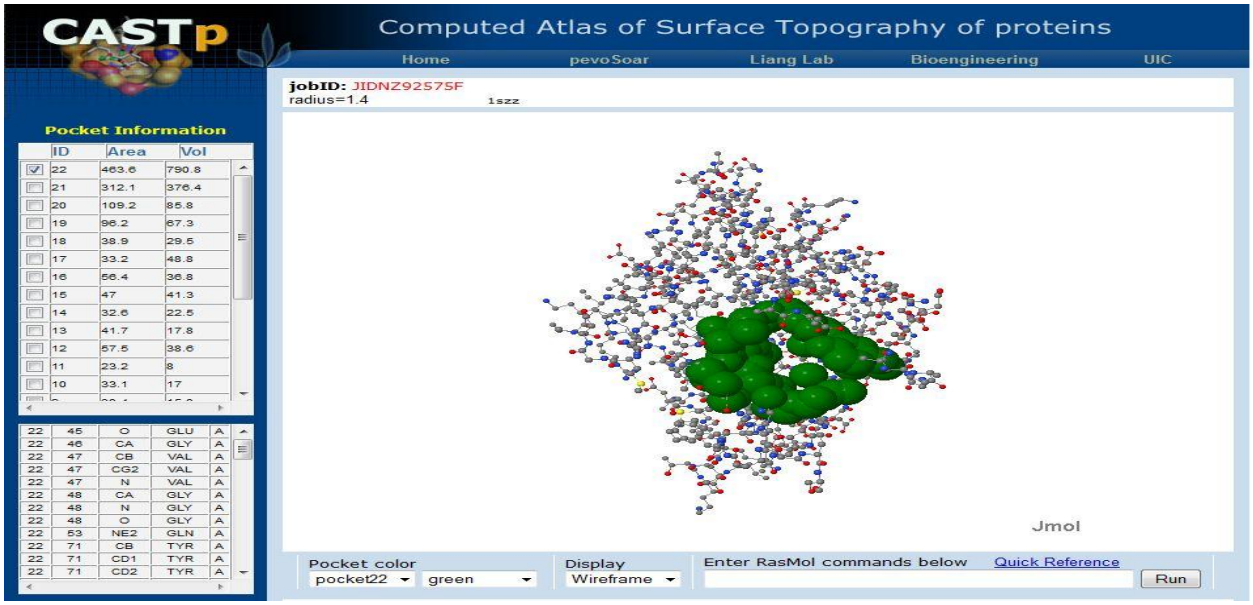


Figure 9: Screenshot of CASTp server interpreting active site of PDF protein.

For more precise docking studies information about active metal binding amino acid residue was obtained from UniProt. The URL of this site is http://www.uniprot.org/uniprot/

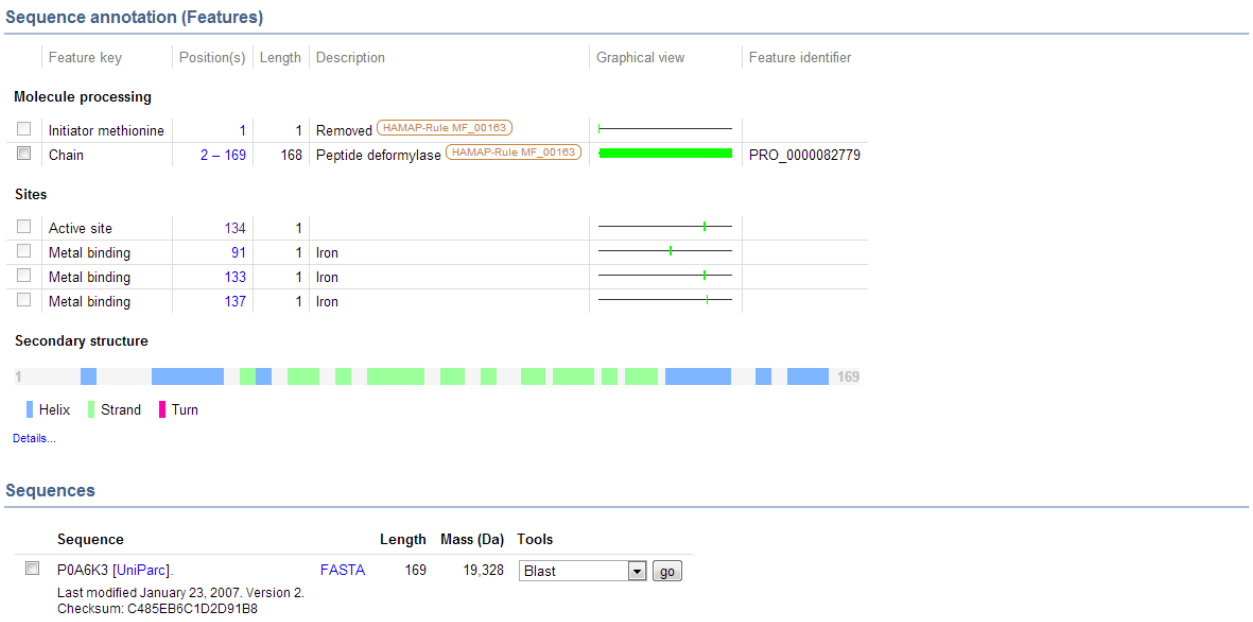


Figure 10: Screenshot of UniProt showing active metal binding amino acid residue.

3.9 Target geometry cleaning and optimization using ArgusLab:

The target obtain from above step is validated in software's where it is converted from higher energy and less stable state to lower energy and more stable state. ArgusLab is a molecular modeling, graphics and drug design software freely available for windows operating system.

- The chosen macromolecule was opened with the help of ArgusLab.
- Water molecules were deleted.
- If extra peptide chain present it was deleted.
- Except the metallic ion other hetero and miscellaneous residues were deleted. Metallic ion was not deleted because it is important for catalytic activity of enzyme.
- Clean the geometry by clicking the clean button on the toolbar the structure of the molecule changes as its geometry is cleaned. This step is followed as long as the macromolecules geometry is not cleaned.

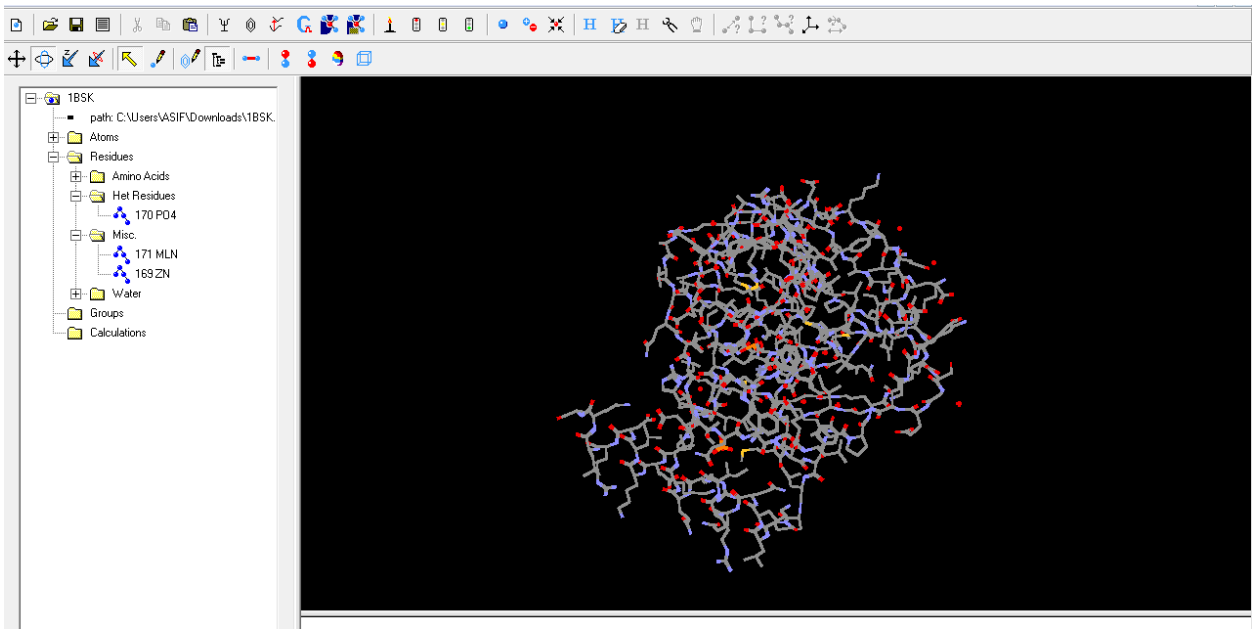


Figure 11: Screenshot of ArgusLab molecular modeling tool showing geometry cleaning of PDF.

3.10 Identification of ligands from PubChem

After going several literature surveys, we found that the molecules which have potential as an antibacterial drug by inhibiting PDF protein were hydroxamic acid derivatives. Actinonin and macrolactin N are the only naturally-occurring PDIs (Peptide Deformylase inhibitors) [37, 68]. Actinonin form the framework for the synthesis, purification, and estimation of more potent PDIs (s), so we choose actinonin as our reference drug. It has already been shows that the hydrazides have a potential to inhibit the PDF's [40]. Taking it in consideration and making our work novel, we proceeded our studied with hydrazine, a related class of hydrazides. We targeted important members of this class called sulfonylhydrazides to study their ability to inhibit Protein Deformylase. The 3D chemical structure of ligands was obtained from PubChem. It is a database which offers information about the biological activities of small molecules and includes substance information, compound structure, bioactivity data, origin and related literatures. Ligand optimization and minimization was done using Argus Lab software. After analyzing their binding energies using Autodock4.2 software we found that compound sulfonylpiperidine gave very good result compared to actinonin. After then eliminating all other molecules we concise our study by focusing on sulfonylpiperidine and searched all derivatives available on PubChem.

We select different derivatives of Sulfonylpiperdine for further study. To get different isomers of sulfonylpipridine first we visited PubChem database webpage; enter “sulfonylpiperidine” in search bar. Selecting the first member of this family we went to structure search tool and search similar compounds with similarity score greater than 90%. It searches 118 different derivatives of sulfonylpiperidine. We accessed PubChem on 21st August, 2012.

3.11 Molecular properties and drug likeness calculation of ligands using Molsoft

Molsoft is web-enabled Cheminformatics system that provides wide range of software tools supporting molecule manipulation and processing, plus SMILES and SD file conversion, standardization of molecules, generation of tautomers, molecule fragmentation, calculation of various molecular properties required in QSAR, molecular modeling and drug design, high quality molecule depiction, molecular database tools supporting substructure and similarity searches, also hold fragment-based virtual screening, bioactivity prediction and data visualization.

- The URL http://molsoft.com/mprop/ was accessed to open Molsoft server.
- In “Enter SMILES” option bar the SMILES format of ligand was paste and software was run to calculate properties and to predict bioactivity.

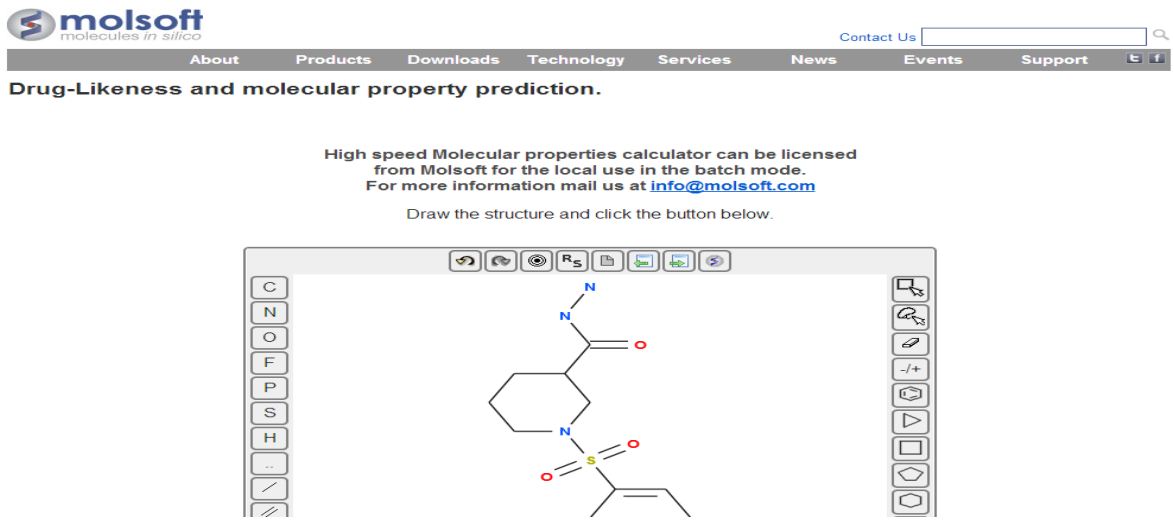
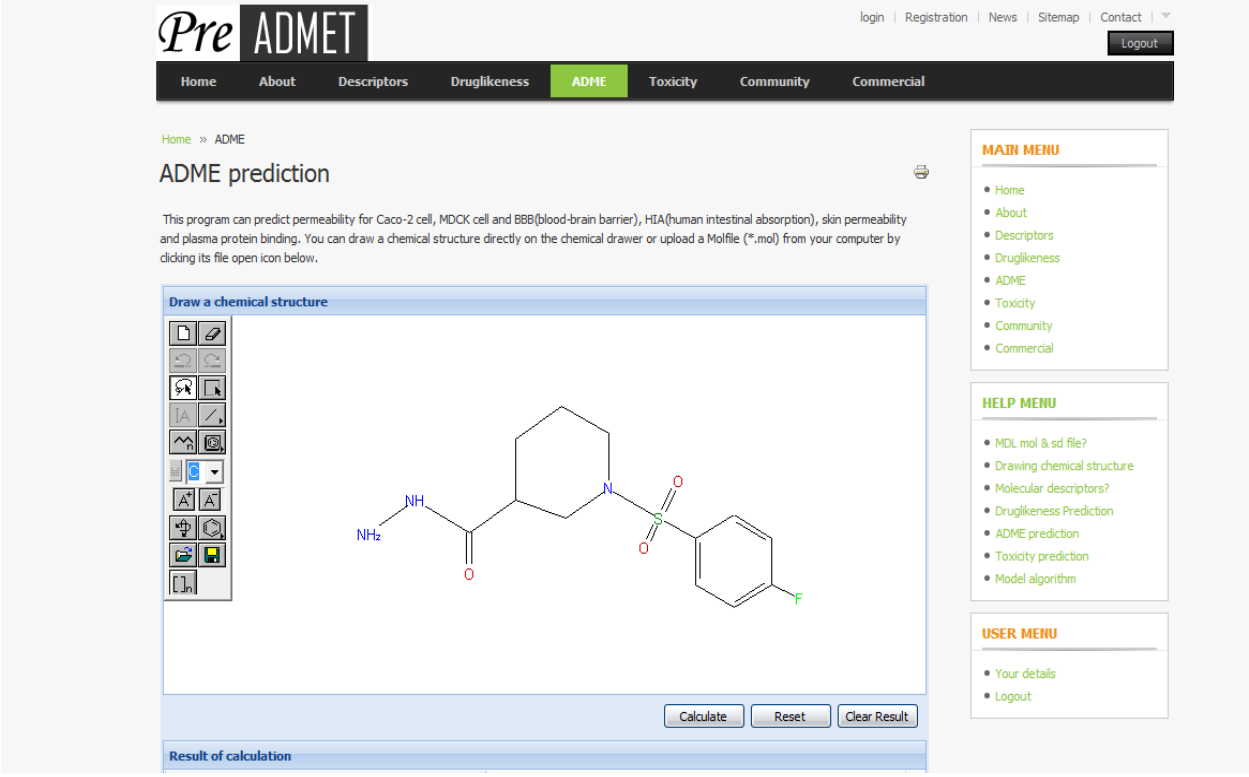


Figure 12: Screenshot of Molsoft molecular properties prediction tool.

3.12 ADME properties and toxicity testing of ligands using PreADMET server

An important tailback remains in the drug discovery process, particular in the later stages of lead discovery, is analysis of the ADME and obvious toxicity properties of drug candidates. More than 50% of the drug candidates failed because of ADME/Toxicity deficiencies during drug development process. So in most pharmaceutical corporations a collection of in vitro ADME/Tox screens has been enforced to avoid this failure at the development of novel drug with the aim of discarding compounds within the discovery section that are seemingly to fail any down the line. PreADMET is a web based application which speedily predicts the drug-likeness and ADME/Tox knowledge of drug candidates.

- The URL http://preadmet.bmdrc.org/ was accessed to open PreADMET.
- Chemical structure of ligand was drew and submitted. The result was saved carefully.



.

Figure 13: Screenshot of PreADMET tool used for ADME/Tox property prediction.

3.13 Energy minimization of all ligand structures by PRODRG server

PRODRG is an online server that takes an account of small molecules as PDB coordinates, MDL molfile/ sdf file/ SYBYL Moll2 file/ text drawing and from it produce a variety of topologies for use with Autodock, GROMACS, REFMAC, SHELX, WHAT IF, HEX, CNS, O and other programs, as well as energy minimized coordinates in a range of formats.

- The URL http://davapc1.bioch.dundee.ac.uk/prodrg/ was accessed to open PRODRG server.
- The token was get by giving email id.
- PubChem was opened, ligand id was entered, sdf file format of ligand was copied and pasted in PRODRG server. The chirality was set to "yes", charges to "full" and energy minimization to "yes".
- Run PRODRG and "pdb (polar hydrogens)" of minimized ligand was saved in pdb format.

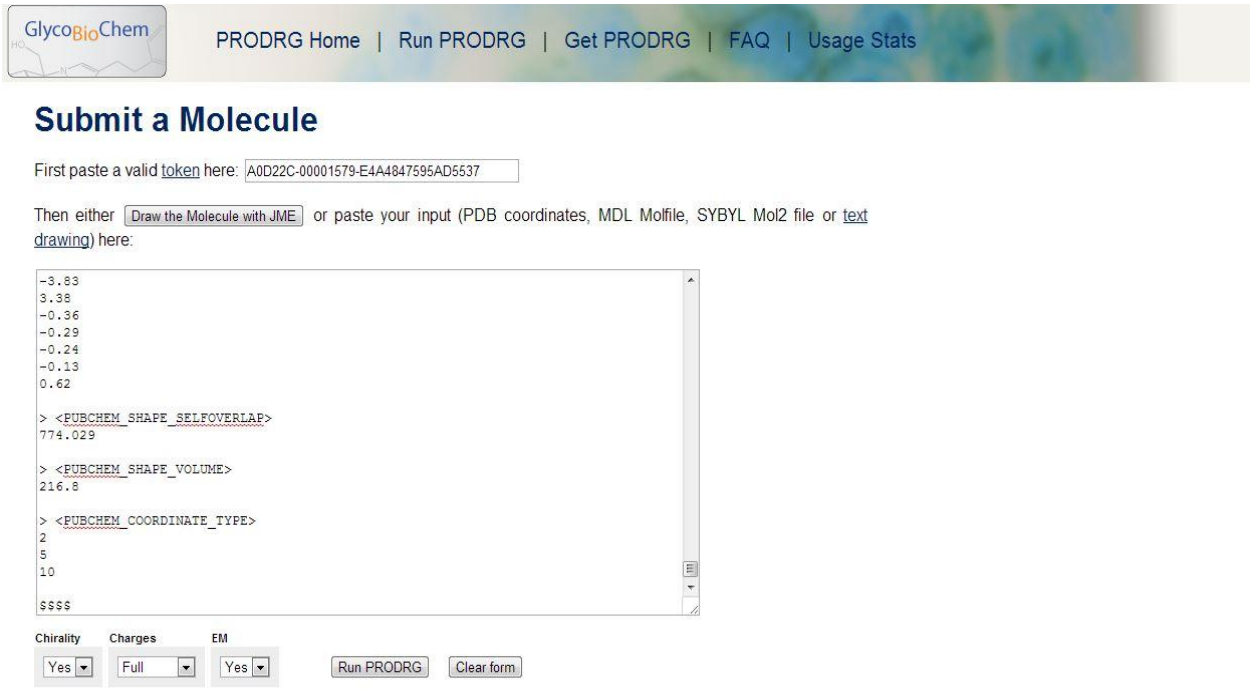


Figure 14: Screenshot of PRODRG energy minimization server.

3.14 Docking of target and ligand by AutoDock4.2 -

AutoDock4.2 is a set of automated tools and is freely available under the GNU General Public License. It is designed to predict how small molecules, like substrates or drug molecules, bind to a receptor ok known 3D structure. Two programs on which AutoDock is based are, autodock program that performs the docking of the ligand to set of grids describing the target protein and the autogrid program that pre calculates these grids. Autodock program employed Lamarckian Genetic Algorithm to explore low binding energy orientations. It is faster than earlier versions and has AMBER force field, free energy scoring function that is based on a regression analysis and a larger set of diverse protein-ligand complexes with known inhibition constants.

AutoDock4.2 used autogrid program to produce grid maps. The spacing between the grids was set default value of 0.375 A^{o}. The centre of grid was set to active metal binding amino acid residue and grid dimensions were set to 60 × 60 × 60 A^{o3}. The standard docking protocol was followed for flexible ligand docking and consisted of 10 independent GA runs for each ligand, using an initial randomly generated population of 150 binding orientations, with 2500000 energy evaluations, maximum number of 27000 generations, gene mutation rate of 0.02 and cross over rate of 0.80. The docked confirmation obtained was further clustered together with accepted RMSD of 0.5 A^{o}. The clusters were ranked by the lowest energy conformation of each cluster. Total 118 ligand molecules were docked against 8 bacterial Peptide Deformylase (PDF) proteins.

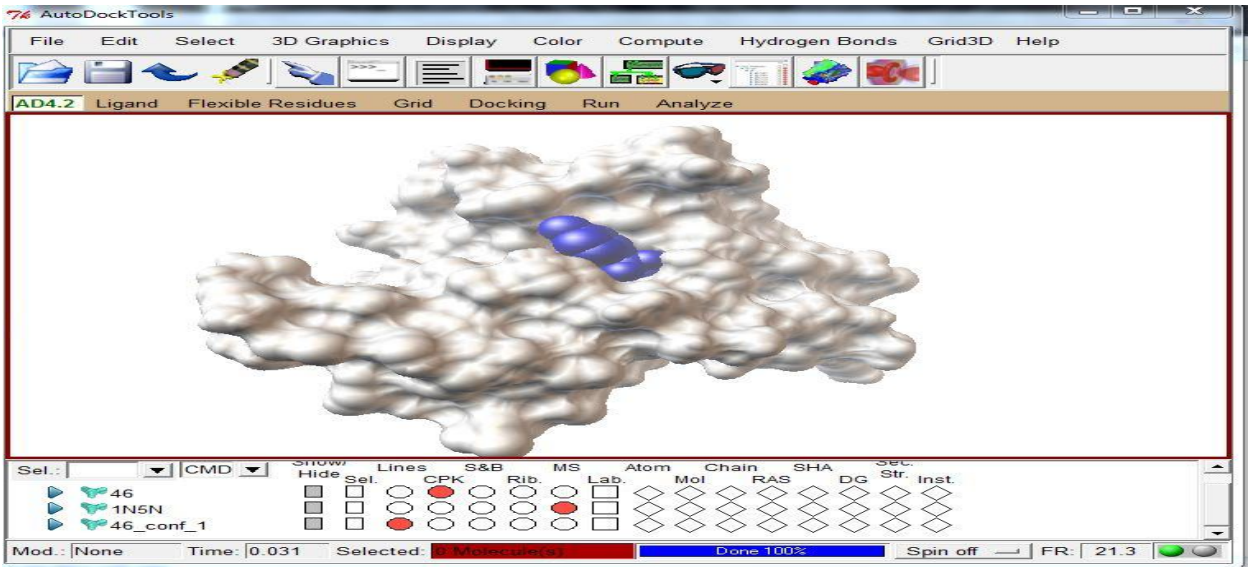


Figure 15: Showing docking of PDF (PDB Id-1N5N) with CID_4539974

3.15 Hydrogen bond analysis of docked complex by Ligplot+ for 2-D and UCSF Chimera for 3-D

LigPlot+ is a freely available, user friendly program that automatically generates schematic diagram of the interaction between protein and ligand for a given PDB file of docked protein-ligand complex and shows these interactions through hydrogen bonds and hydrophobic contacts. The H-bonds are represented by dashed lines between the atoms of ligand and protein, whereas, hydrophobic contacts are indicated by a convex arc, facing towards the ligand and radiating spikes. The contacted atoms are represented by concave arc towards each other and spoke radiating back.

UCSF Chimera is a freely available on web and is highly extensible program. It is used for visualization and analysis of molecular structures and related data, including docking results, trajectories, conformational ensembles, density maps, sequence alignment and supramolecular assemblies. It generates images and animation of structures.

We analyzed the docked structures with these two softwares. LigPlot+ is used for 2-D visualization and UCSF Chimera for 3-D visualization of Hydrogen bonds formed between protein and ligand molecule. UCSF Chimera represents the H-bond by blue and orange lines.

3.16 Pharmacophore modeling of best docked complex by LigandScout3.1

Pharmacophore models representing ligand-receptor interaction are derived from interaction of each molecule of ligand with its surrounding amino acid residues [69] and represents collection of molecular features which are necessary for molecular recognition and interaction of ligand by a protein. Pharmacophore modelling is playing a key role in the identification of ligand features for the particular targets [70-71].

We have done pharmacophore modeling using LigandScout3.1 software. It takes docked protein-ligand complex and identifies the key features on the ligand. That features help the ligand to interact with points on the protein.

3.17 Molecular dynamics simulation studies of best docked complex by GROMACS4.05

For MD simulation GROMACS (GROningen Machine for Chemical Simulations) software was used. It is a freely available and open source molecular dynamics simulation package. It is extremely fast and mainly designed for simulations of biochemical molecules like proteins, lipids and nucleic acids. It is operated via command-line and use files for input and output.

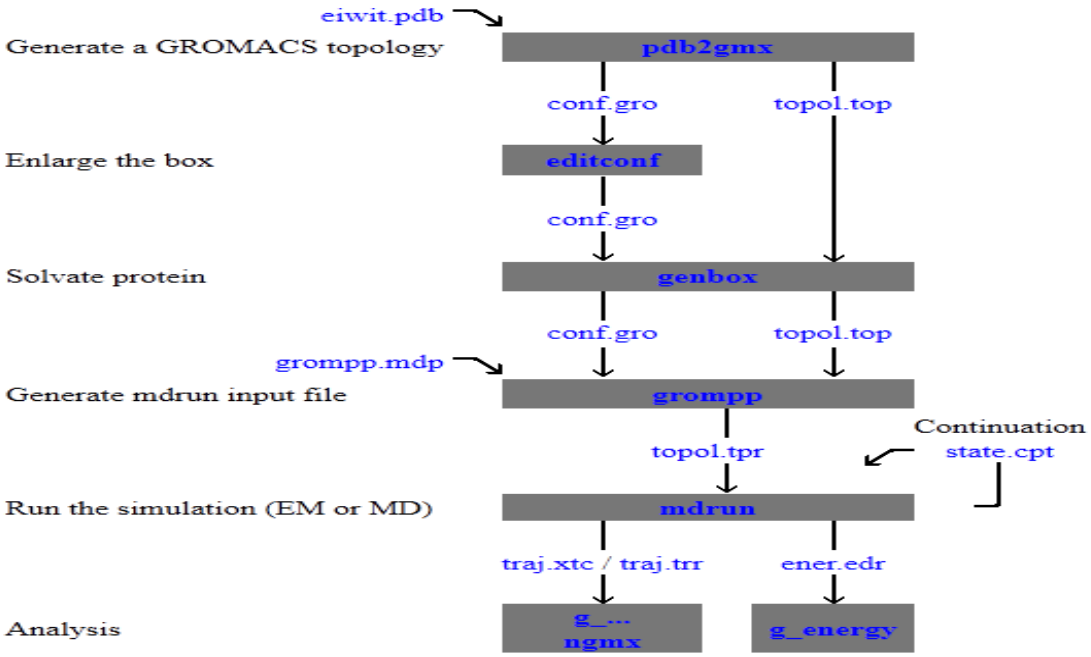


Figure 16: Showing the basic molecular dynamics simulation strategy by Gromacs tool

The molecular dynamics simulation was performed with GROMACS version 4.05 software packages for the molecules giving best docking and H-bond analysis result. The forcefield applied was GROMOS4031. The initial structure of protein was placed in a cubic box and filled with water. The energy minimization of the system was performed without restraint using steepest descent method for 100 steps. Then molecular dynamics equilibration was run at 300 K, with decreasing harmonic position restraints on the protein atoms over a 20 ps time interval, followed by 1 ns of MD equilibration without restraints. The P-LINCS algorithm was used for MD simulation. A dielectric constant of 1 and a time step of 2 fs were used. The temperature and pressure was kept constant during the MD simulations.

CHAPTER 4

RESULT AND DISCUSSION

4. RESULT AND DISCUSSION

4.1 Result of Multiple Sequence Alignment

The multiple sequence alignment by Clustal Omega showed that the active site of PDF is conserved throughout the microorganisms, composing of three sequence motifs, motif 1 [GφGLAAXQ], motif 2 [EGCLS], and motif 3 [QHEφDH] (where φis a hydrophobic amino acid and X is any amino acid) (where φis a hydrophobic amino acid and X is any amino acid). Therefore, PDF appears to be a broad spectrum novel antibacterial drug target.

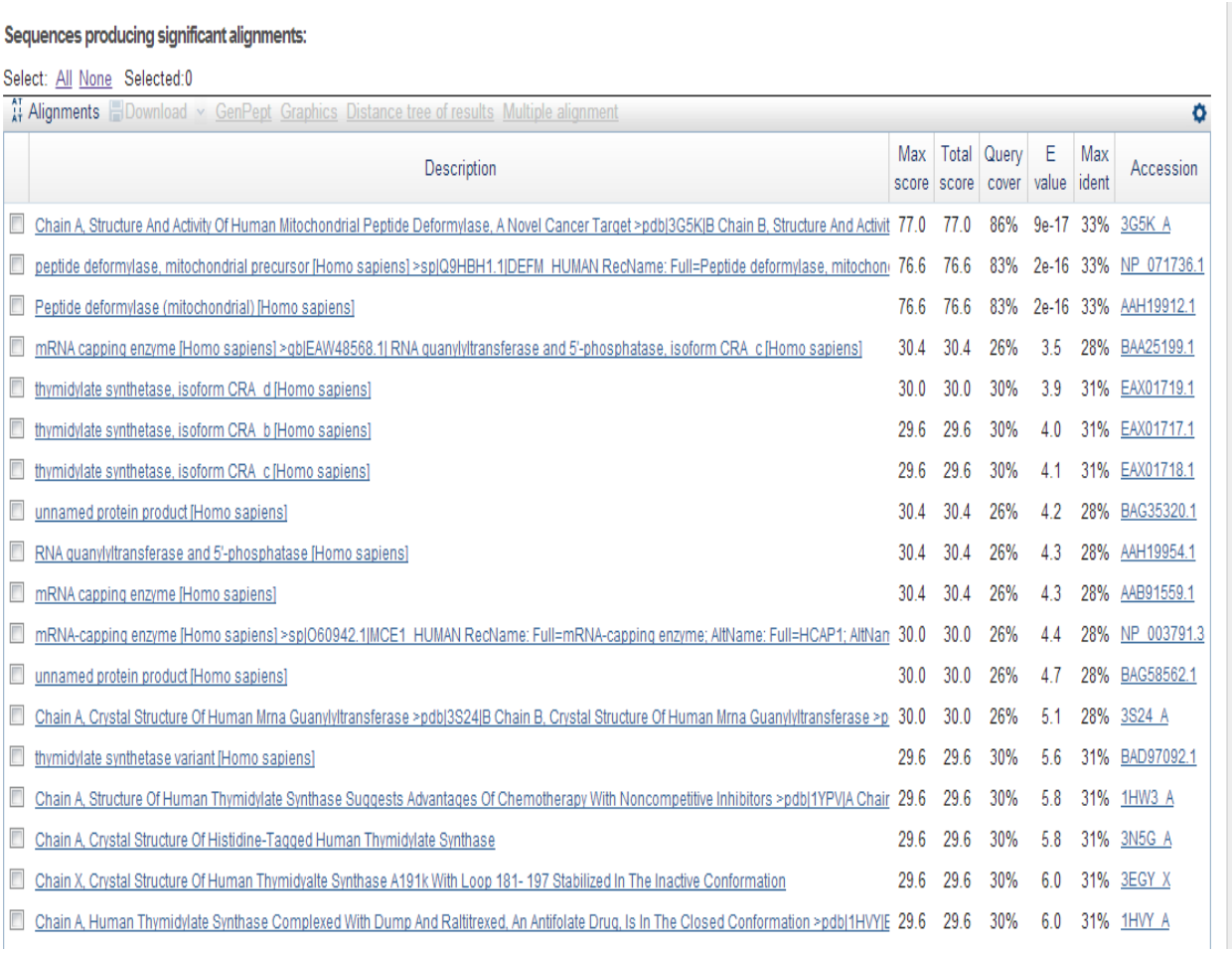
```
gi|75765204|pdb|1SZZ|       LRKISEPVTED----EIQT-KEF----KKLIRDMFTTMRH-AEGVGLAAPQIGILKQIVV
gi|7767089|pdb|1BSK|        LRKVAKPVEEV----N----AEI----QRIVDDMFETMYA-EEGIGLAATQVDIHQRIIV
gi|224510906|pdb|3G6N|      LREVAKEVSLPLSEEDISLGKEMLEFLKNSQDPIKAEELHLRGGVGLAAPQLDISKRIIA
gi|325534093|pdb|3QU1|      LRVQSKQVTDV----A-----SV----QTLIDDLLDTLYATDNGIGLAAPQVGREEAIVV
gi|33357406|pdb|1N5N|       LRTIAKPVEVV----D----DAV----RQLIDDMFETMYE-APGIGLAATQVNVHKRIVV
gi|22219285|pdb|1LQW|       LRQKAAELELPLTKEEKETLIAMREFLVNSQDEEIAKRYGLRSGVGLAAPQINISKRMIA
gi|62738425|pdb|1XEN|       LRKVAKPVEEV----N----AEI----QRIVDDMFETMYA-EEGIGLAATQVDIHQRIIV
gi|316983220|pdb|3L87|      LRAVAQDVTFPLNEDDIILGEKMLQFLKNSQDPVTAEKMELRGGVGLAAPQLDISKRIIA
                            **  :  :              .                    *:**** *:   : ::.

gi|75765204|pdb|1SZZ|       VGSEDNERYPGTPD----VPERIILNPVITPLTKD--TSGFWEGCLSVPG-MRGYVERPN
gi|7767089|pdb|1BSK|        IDVSENRDE-----------RLVLINPELLEKSGE---TGIEEGCLSIPE-QRALVPRAE
gi|224510906|pdb|3G6N|      VHVPSPDPEADG-----PSLSTVMYNPKILSHSVQDACLGEGEGCLSVDREVPGYVVRHA
gi|325534093|pdb|3QU1|      IDLSDNRDQ-----------PLVLINPKVVSGSNK---EMGQEGCLSVPD-YYADVERYT
gi|33357406|pdb|1N5N|       MDLSEDKSE-----------PRVFINPEFEPLTEE--MDQYQEGCLSVPG-FYENVDRPQ
gi|22219285|pdb|1LQW|       VLIPDDGSG--------KSYDYMLVNPKIVSHSVQEAYLPTGEGCLSVDDNVAGLVHRHN
gi|62738425|pdb|1XEN|       IDVSENRDE-----------RLVLINPELLEKSGE---TGIEEGCLSIPE-QRALVPRAE
gi|316983220|pdb|3L87|      VLIPNPEDKDGNPPKEAYALKEVMYNPRIIAHSVQDAALADGEGXLSVDRVVEGYVIRHS
                            :   .                :: ** .   : .       ** **:       * *

gi|75765204|pdb|1SZZ|       QIRMQWMDEKGNQFDETIDGYKAIVYQHECDHLQGILYVDRLKDTKLFGFNETLDSSHNV
gi|7767089|pdb|1BSK|        KVKIRALDRDGKPFELEADGLLAICIQHEMDHLVGKLFMDYLSPLKQQRIRQKVEKLDRL
gi|224510906|pdb|3G6N|      KITVSYYDMNGEKHKIRLKNYESIVVQHEIDHINGVMFYDHINDQNPFALKEGVLVIE--
gi|325534093|pdb|3QU1|      SVVVEALDREGKPLRIETSDFLAIVMQHEIDHLSGNLFIDYLSPLKQQMAMKKVKKHVKN
gi|33357406|pdb|1N5N|       KVRIKALDRDGNPFEEVAEGLLAVCIQHECDHLNGKLFVDYLSTLKRDRIRKKLEKQHRQ
gi|22219285|pdb|1LQW|       RITIKAKDIEGNDIQLRLKGYPAIVFQHEIDHLNGVMFYDHIDKNHPLQPHTDAVEV---
gi|62738425|pdb|1XEN|       KVKIRALDRDGKPFELEADGLLAICIQHEMDHLVGKLFMDYLSPLKQQRIRQKVEKLDRL
gi|316983220|pdb|3L87|      RVTIEYYDKNSDKKKLKLKGYQSIVVQHEIDHTNGIMFFDRINEKNPFEIKEGLLLIE--
                             : :   * ...      .   ::  *** **  * :: * :.  .
```

Figure 17: Clustal Omega Multiple Sequence Alignment analysis of bacterial PDF protein sequences.

4.2 BLAST output

The BLAST of bacterial Peptide Deformylase protein sequence against non redundant protein sequences of *Homo sapiens* (taxid: 9606) by pBLAST program showed similarity below 40% and confirmed the absence of equivalent functional PDF protein in human. The essentiality of

PDF protein for bacterial protein synthesis along with conserved active site region in bacterial population and the absence of functionally equivalent protein in human fulfil the"essentiality" and "selectivity" criteria for choosing PDF as a novel drug target.



Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Chain A, Structure And Activity Of Human Mitochondrial Peptide Deformylase, A Novel Cancer Target >pdb|3G5K|B Chain B, Structure And Activit	77.0	77.0	86%	9e-17	33%	3G5K_A
peptide deformylase, mitochondrial precursor [Homo sapiens] >sp|Q9HBH1.1|DEFM_HUMAN RecName: Full=Peptide deformylase, mitochon	76.6	76.6	83%	2e-16	33%	NP_071736.1
Peptide deformylase (mitochondrial) [Homo sapiens]	76.6	76.6	83%	2e-16	33%	AAH19912.1
mRNA capping enzyme [Homo sapiens] >gb|EAW48568.1| RNA guanylyltransferase and 5'-phosphatase, isoform CRA_c [Homo sapiens]	30.4	30.4	26%	3.5	28%	BAA25199.1
thymidylate synthetase, isoform CRA_d [Homo sapiens]	30.0	30.0	30%	3.9	31%	EAX01719.1
thymidylate synthetase, isoform CRA_b [Homo sapiens]	29.6	29.6	30%	4.0	31%	EAX01717.1
thymidylate synthetase, isoform CRA_c [Homo sapiens]	29.6	29.6	30%	4.1	31%	EAX01718.1
unnamed protein product [Homo sapiens]	30.4	30.4	26%	4.2	28%	BAG35320.1
RNA guanylyltransferase and 5'-phosphatase [Homo sapiens]	30.4	30.4	26%	4.3	28%	AAH19954.1
mRNA capping enzyme [Homo sapiens]	30.4	30.4	26%	4.3	28%	AAB91559.1
mRNA-capping enzyme [Homo sapiens] >sp|O60942.1|MCE1_HUMAN RecName: Full=mRNA-capping enzyme; AltName: Full=HCAP1; AltNan	30.0	30.0	26%	4.4	28%	NP_003791.3
unnamed protein product [Homo sapiens]	30.0	30.0	26%	4.7	28%	BAG58562.1
Chain A, Crystal Structure Of Human Mrna Guanylyltransferase >pdb|3S24|B Chain B, Crystal Structure Of Human Mrna Guanylyltransferase >p	30.0	30.0	26%	5.1	28%	3S24_A
thymidylate synthetase variant [Homo sapiens]	29.6	29.6	30%	5.6	31%	BAD97092.1
Chain A, Structure Of Human Thymidylate Synthase Suggests Advantages Of Chemotherapy With Noncompetitive Inhibitors >pdb|1YPV|A Chair	29.6	29.6	30%	5.8	31%	1HW3_A
Chain A, Crystal Structure Of Histidine-Tagged Human Thymidylate Synthase	29.6	29.6	30%	5.8	31%	3N5G_A
Chain X, Crystal Structure Of Human Thymidyalte Synthase A191k With Loop 181- 197 Stabilized In The Inactive Conformation	29.6	29.6	30%	6.0	31%	3EGY_X
Chain A, Human Thymidylate Synthase Complexed With Dump And Raltitrexed, An Antifolate Drug, Is In The Closed Conformation >pdb|1HVY|E	29.6	29.6	30%	6.0	31%	1HVY_A

Figure 18: BLAST analysis of bacterial PDF protein sequence against non redundant human protein sequences.

4.3 Active site prediction

Active sites of the target PDF protein of all organisms were predicted by CASTp, active site prediction tool. CASTp computation is based on the pocket algorithm of the alfa shape theory. The active site of protein with structural pockets and cavities were calculated for all bacteria's. The result is shown below:

Table 2: CASTp calculation on active site residues.

S.No.	Protein Model	Active Site Residues
1	1BSK	Val5, His7, Glu41, Glu42,Gly43, Ile44, Gly45, Leu46, Gln50, Ile86, Glu 87, Glu88, Gly89, Cys90, Leu91, Ile93, Pro94, Glu95, Gln96, Arg97, Leu125, Ile128, Cys129, His132, Glu133, His136, Leu161, Asp162, Lys165
2	1LQW	Ile77, Pro78, Asp80, Ser82, Gly83, Lys84, Tyr86, Tyr88, Cys111, Arg127, Ile128, Gln141, Leu142, Arg143, Leu144, Lys145, Gly146, Tyr147, Pro148,
3	1N5N	Met39, Tyr40, Pro43, Gly44, Ile45, Gly46, Gln51, Asp62, Ser64, Glu65, Lys67, Tyr88, Gln89, Glu90, Gly91, Cys92, Leu93, Val95, Pro96, Gly97, Phe98, Tyr99, Leu127, Val130, Cys131, His134, Glu135, His138
4	1XEN	Glu41, Glu42, Gly43, Ile44, Gly45, Gln50, Ile86, Glu87, Glu88, Gly89, Cys90, Leu91, Ile93, Pro94, Glu95, Gln96, Arg97, Leu125, Ile128, Cys129, His132, Glu133, His136
5	3G6N	Arg56, Gly57, Gly58, Val59, Gly60, Gln65, Leu108, Glu112, Gly113, Cys114, Leu115, Tyr150, Ile153, Val154, His157, Glu158, His161
6	3L87	Arg68, Gly70, Val71, Gly72, Leu73, Gln77, Leu125, Glu129, Gly130, Leu132, Ser133, Val134, Arg136, Arg144, Tyr167, Ile170, Val171, His174, Glu175, Asp177, His178, Ile182, Met183, Phe184,
7	3QU1	Met1, Ala2, Val3, Leu4, Glu5, Ile6, Leu7, Thr8, Thr37, Leu38, Tyr39, Ala40, Thr41, Asp42, Asn43, Gly44, Ile45, Asp62, Leu63, Ser64, Aso65, Asn66, Arg67, Gly90, Cys91, Leu92, Val94, Pro95, Asp96, Tyr97, Tyr98, Phe126, His162, Val163, Arg166, Arg168
8	1SZZ	Glu45, Gly46, Val47, Gly48, Gln53, Tyr71, Thr74, Phe97, Trp98, Glu99, Gly100, Cys101, Leu102, Val104, Pro105, Gly106, Met107, Arg108, Tyr136, Ile139, Val140, His143, Glu144, Asn166

Furthermore, amino acid residues that involved in bonding with metal residue of protein i.e. active amino acid residues were obtained from UniProt server.

Table 3: UniProt analysis of active metal binding amino acid residues.

PDB id	1N5N	1BSK	3G6N	1SZZ	3L87	3QU1	1LQW	1XEN
UniProt Active amino acid residue	Cys92	Cys90	Cys114	Cys101	Cys131	Cys91	Cys111	Cys90

4.4 Docking result

The docking of protein with ligand molecules was done step by step according to the standard procedure. It was done by automated docking tool AutoDock4.2. In order to find the novel drug against PDF first we docked the PDF enzyme of all selected bacteria with reference molecule. The reference molecule was selected Actinonin (CID_443600) which is a natural PDI (Peptide Deformylase Inhibitor). For most of the ligands, docking explicitly generated the crystallographic binding orientation within the protein cavity. For each ligand, 10 independent docking run were initiated with randomized populations and docking results for individual run were clustered if their final docked positions were within a tolerance of 0.5 A^0.

The docking result of PDF enzymes of selected bacteria with reference molecule actinonin is listed below:

Table 4: Docking results of PDF with actinonin (CID_443600)

S.No.	Organism	PDB iD	Binding energy (-kcal/mol)	Ki value (µmol)
1	*Pseudomonas aeruginosa*	1N5N	7.66	2.44
2	*Esherichia coli*	1BSK	7.23	4.99
3	*Enterococcus faecium*	3G6N	6.61	13.1
4	*Leptospira interrogans*	1SZZ	7.38	9.35
5	*Sreptococcus mutans*	3L87	6.01	39.24
6	*Vibrio cholerae*	3QU1	7.55	2.93
7	*Staphylococcus aureus*	1LQW	6.96	7.85
8	*Esherichia coli*	1XEN	7.07	6.57

Here we see that the highest binding affinity of actinonin is -7.66 kcal/mol with inhibition constant of 2.44 µmol. Actinonin gave this best result with PDF of *Pseudomonas aeruginosa* (1N5N).

The top five best docking results of different derivatives of sulfonylpiperidine with PDF of all selected microorganism is given below:

Table 5: Top five best docking results of isomers of sulfonylpiperidine with PDF

S.No	Organism	PDB Id	Compound Id	Binding energy (-kcal/mol)	Ki (nmol)
1	*Pseudomonas aeruginosa*	1N5N	CID_3642762**	9.98	48.3
			CID_6391521	9.77	68.76
			CID_4025857	9.53	102.57
			CID_4538988	9.45	117.41
			CID_4539974*	9.24	167.54
2	*Esherichia coli*	1BSK	CID_4539974*	10.54	18.74
			CID_4025857	9.65	84.57
			CID_2761095	9.38	132.81
			CID_3642762**	9.28	157.64
			CID_2761072	9.11	209.1
3	*Enterococcus faecium*	3G6N	CID_4539974*	9.67	81.02
			CID_3642860	9.2	181.44
			CID_3314532	8.87	314.55
			CID_4268983**	8.82	340.78
			CID_3642762**	8.75	385.56
4	*Leptospira interrogans*	1SZZ	CID_4539974*	9.68	80.65
			CID_3642860	9.3	152.77
			CID_4268983**	9.02	244.85
			CID_3642762**	8.76	377.45
			CID_581405	8.64	464.43
5	*Sreptococcus mutans*	3L87	CID_3314532	9.04	234.99
			CID_300664	8.76	376.04
			CID_4539974*	8.75	383.02

			CID_2886671	8.72	406.42
			CID_4268983**	8.61	485.42
6	*Vibrio cholerae*	3QU1	CID_6391521	10.42	23.13
			CID_3642762**	10.31	27.92
			CID_5068042	10.25	30.59
			CID_4539974*	9.91	54.11
			CID_4268983**	9.8	65.61
7	*Staphylococcus aureus*	1LQW	CID_3314532	10.36	25.42
			CID_4539974*	9.79	66.22
			CID_3642860	9.48	112.56
			CID_4268983**	9.16	191.76
			CID_5068042	9.14	199.53
8	*Escherichia coli*	1XEN	CID_4025857	9.79	67.06
			CID_4539974*	9.71	76.87
			CID_3642860	9.48	112.29
			CID_3642762**	9.35	139.85
			CID_4268983**	9.09	216.83

From all docking results we sort out top five docking result for each bacterial PDF-sulfonylpiperidine complex and represent in tabulated form above. We have sorted this result according to decreasing order of binding energy. Comparing from actinonin, we find that even the highest binding energy of sulfonylpiperidine -8.61 kcal/mol (CID_4268983- 3L87 complex) is lower than the lowest binding energy of actinonin -7.66 (CID_443600- 1N5N complex). This result concluded that the sulfonylpiperidine has better inhibition capability for PDF than actinonin. Moreover we see that, the molecule CID_4539974 (shown with *) is present in top five best docking result of all microorganism. Molecules CID_3642762 and CID_4268983 (shown with **) are present in top five best docking result of six microorganism. Based on this we filtered our result by selecting these three molecules CID_4539974, CID_4268983 and CID_3642762 for further analysis, narrowing our path with aim to get a novel potent Peptide Deformylase Inhibitor (PDI).

4.5 Hydrogen bond analysis of best docked molecules

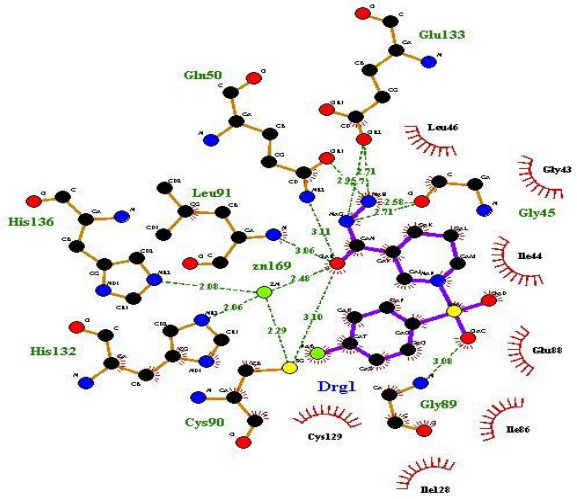
The residues involved in hydrogen bonding with drug candidate were analysed. LigPlot+ was used for 2-D and UCSF Chimera for 3-D analysis and confirmation of LigPlot+ results. The analysis was done for top 3 best docked structures. The result is shown below:

Table 6: Amino acid residues to which ligand made H-bond and number of H-bonds made.

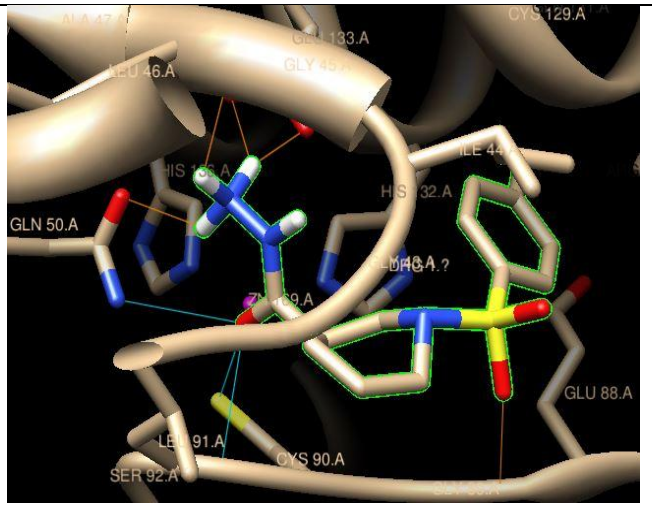
S.No	Organism	PDB Id	Active AA residue	Compound Id	No. of H-bonds	Hydrogen bonds with amino acid residue
1	P.aeruginosa	1N5N	**Cys92**	CID_4539974	6	Gly46,Tyr88,Gly91,**Cys92,** Cys 131, His134
				CID_3642762	3	**Cys92**, Leu93, Gly97
				CID_4268983	1	Gly97
2	E.coli	1BSK	**Cys90**	CID_4539974	8	Gly45, Gln50 (2), Gly89, **Cys90**, Leu91, Glu 133 (2)
				CID_3642762	3	Ile44, Gln50, Glu133
				CID_4268983	4	Gly45, Gln50, Glu95, Glu 133
3	E.faecium	3G6N	**Cys114**	CID_4539974	6	Gly60, Gln65, Gly 113, Tyr150, Glu 158 (2)
				CID_3642762	4	Gln65, Gly113, Leu115, Tyr150
				CID_4268983	2	Gly113, Val154
4	L.interrogans	1SZZ	**Cys 101**	CID_4539974	5	Gln53 (2), **Cys101**, Leu102, His143
				CID_3642762	5	Gln53,Gly100, **Cys101** (2), Leu102
				CID_4268983	7	Val47,Gly48, Gln53 (2),Tyr71, **Cys101**, Leu102
5	S.mutans	3L87	Cys131 **Leu132**	CID_4539974	4	Gln77, **Leu132**, Glu177 (2)
				CID_3642762	4	Gln77, **Leu132**, Gly130 (2)
				CID_4268983	4	Val71, Gln77,**Leu 132**, Glu175
6	V.cholerae	3QU1	**Cys91**	CID_4539974	6	Gly46, Gln51, Gly90, **Cys91**, Glu134, His137
				CID_3642762	5	Ile45, Gly46, **Cys91**, Leu92, Asp96
				CID_4268983	4	Leu92, Gln88, Glu134, His137

7	S.aureus	1LQW	**Cys111**	CID_4539974	4	Gln65, **Cys111**, Leu112, His158
				CID_3642762	3	Gln65, Gly110, Leu114
				CID_4268983	4	Gln65, Gly110, Leu112, Glu155
8	E.coli	1XEN	**Cys90**	CID_4539974	3	**Cys90**, Leu91, Glu133
				CID_3642762	5	Ile44, Gly45, **Cys90**, Leu91, Glu95
				CID_4268983	4	Leu91, Arg97, Glu133, His136

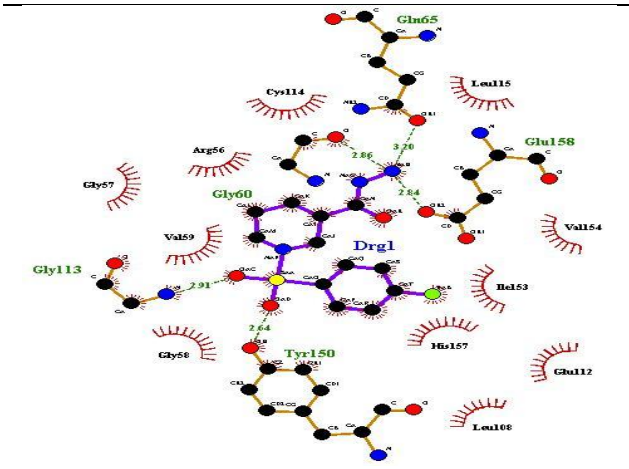
The pictorial representation of of best docked molecule is in figure 19. In the LigPlot+ analysis the dotted line shows the hydrogen bonding between the amino acid residues of protein and drug candidate. The violet coloured molecule is drug candidate. In the UCSF Chimera analysis the blue and orange coloured line shows hydrogen bonding and the fluorescent molecule is the drug candidate. The analysis of next two docked molecule were done in the same way and result shown in above table.



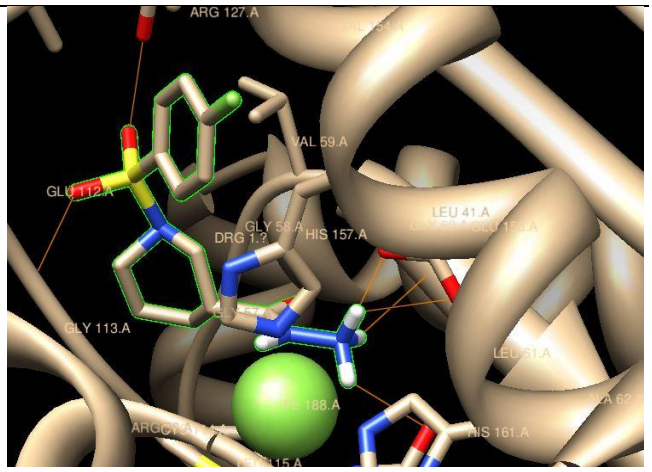
1BSK-CID_4539974 ligplot analysis



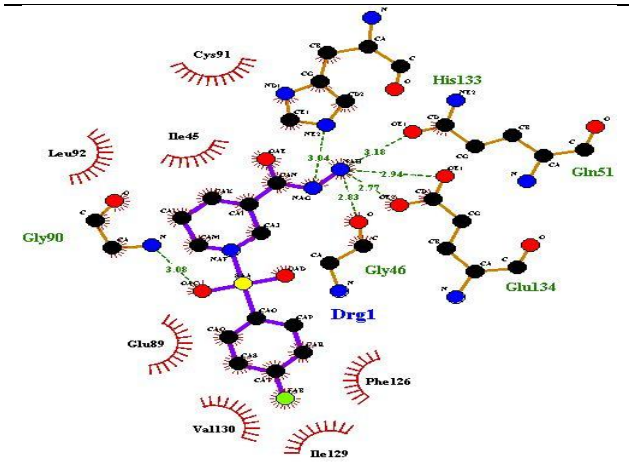
1BSK-CID_4539974 UCSF chimera analysis



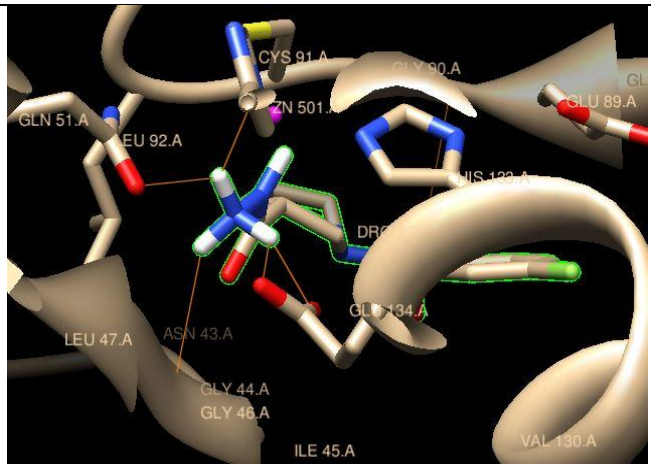
3G6N-CID_4539974 ligplot analysis



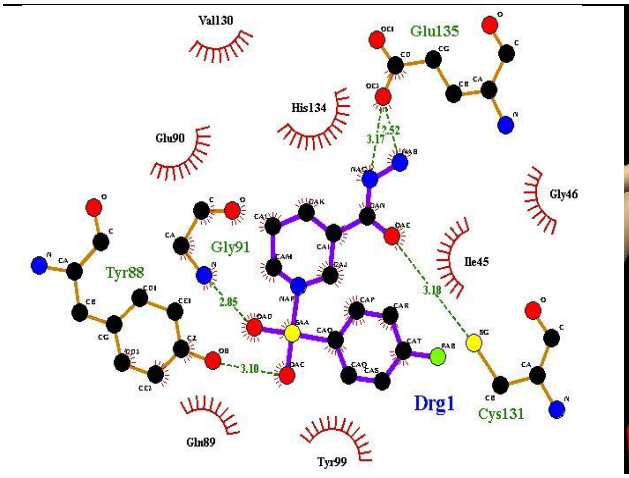
3G6N-CID_4539974 UCSF chimera analysis



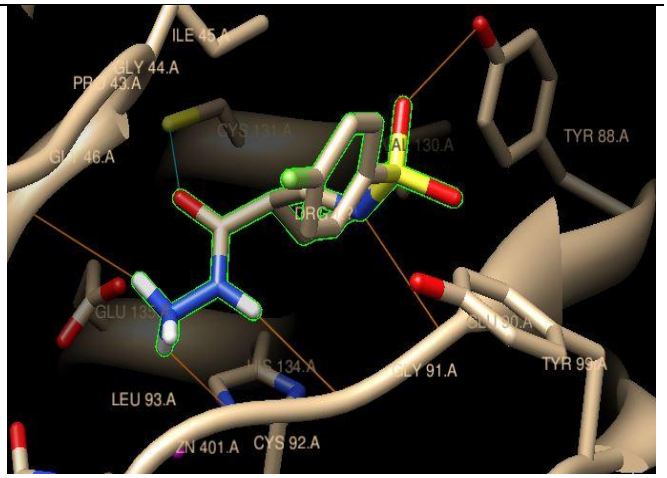
3QU1-CID_4539974 ligplot analysis



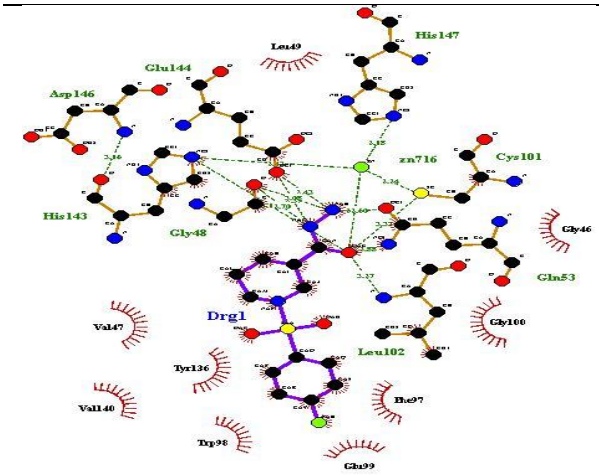
3QU1-CID_4539974 UCSF chimera analysis



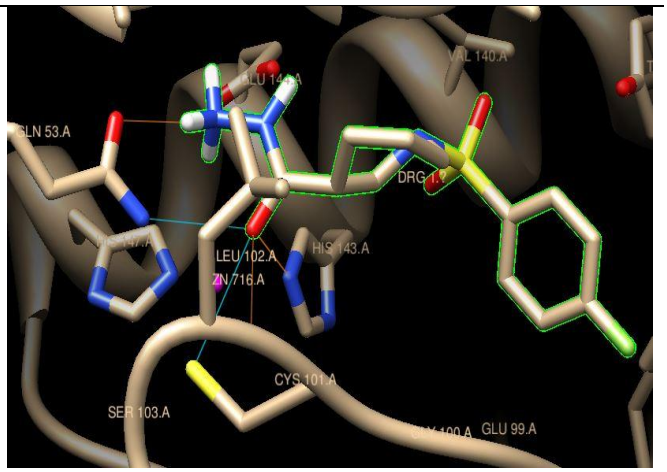
1N5N-CID_4539974 ligplot analysis



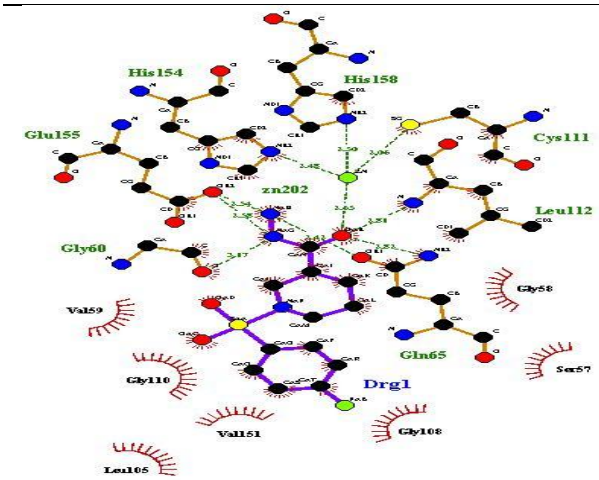
1N5N-CID_4539974 UCSF chimera analysis



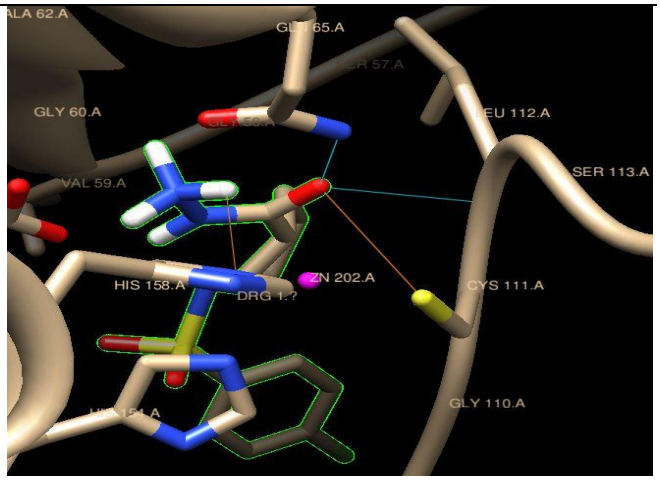
1SZZ-CID_4539974 ligplot analysis



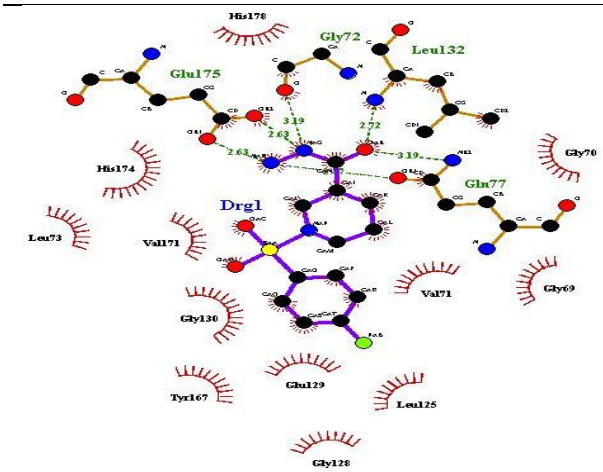
1SZZ-CID_4539974 UCSF chimera analysis



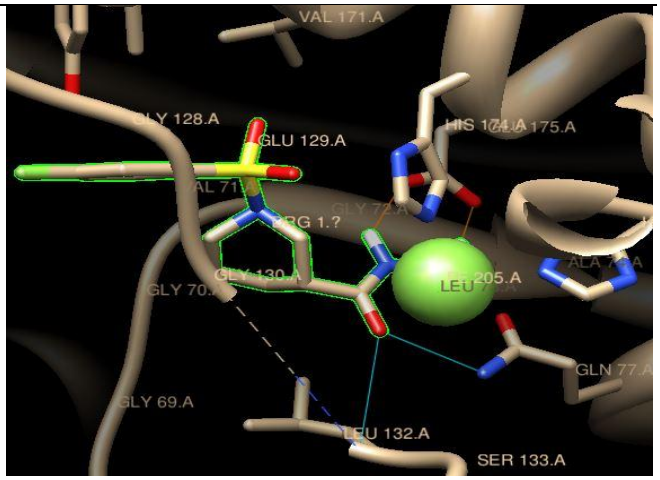
1LQW-CID_4539974 ligplot analysis



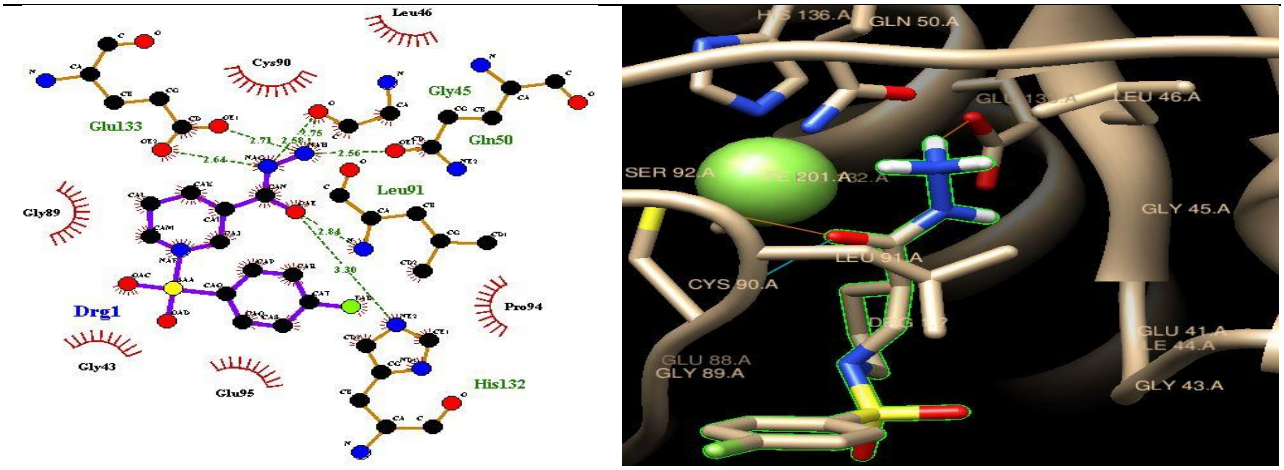
1LQW-CID_4539974 UCSF chimera analysis



3L87-CID_4539974 ligplot analysis



3L87-CID_4539974 UCSF chimera analysis



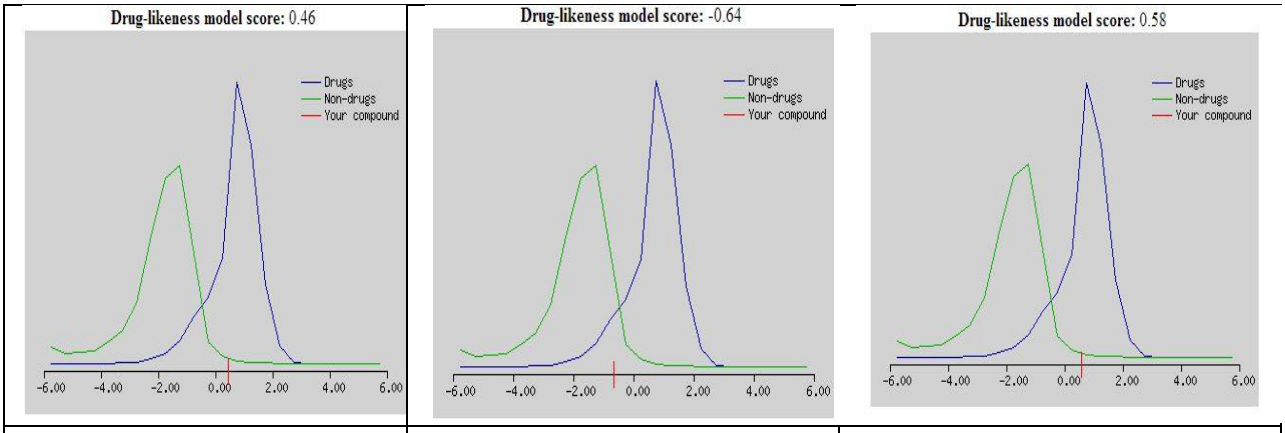
1XEN-CID_4539974 ligplot analysis | 1XEN-CID_4539974 UCSF chimera analysis

Figure 19: Hydrogen bond analysis of CID_4539974 by LigPlot+ and UCSF Chimera

The results of table show hydrogen bonding between drug candidate and amino acid residues of target protein. Furthermore, from the UniProt we identified the active amino acid residue for each protein that is bonded with metal residue. The active metal binding amino acid residue is crucial for the function of protein. This residue directly participates in catalysis and if we hinder this residue to make bond with metal ion, the biological functioning of protein is totally busted. This will add plus point to docking result. The position of active metal binding residue in the protein sequence is also shown in the table. From the result we see that the molecule 1-[(4-methoxyphenyl)sulfonyl]piperidine-4-carbohydrazide (CID_4268983) makes hydrogen bond with active metal binding amino acid residue of 1SZZ and 3L87, molecule 1-[3-(tetrazol-1-yl)phenyl]sulfonylpiperidine (CID_3642762) makes H-bond with active metal binding amino acid residue of 1N5N, 1SZZ, 3L87 and 1XEN and molecule 1-(4-fluorophenyl)sulfonylpiperidine-3-carbohydrazide (CID_4539974) makes H-bond with active metal binding amino acid residue of almost all protein except 3G6N. In case of 3L87, the crystallized protein missed of Cys131 residue, so we consider its neighbor residue Leu132 as centre molecule.

4.6 Molecular, ADME properties calculation and toxicity testing result:

The molecular properties calculation of 3 best docked molecules was done by molsoft online server, the ADME properties calculation and the toxicity testing was done by PreADMET server. The Molsoft calculates the molecular weight, number of hydrogen bond donor, number of hydrogen bond acceptor, octanol-water partition coefficient (LogP), and solubility (LogS). The PreADMET calculates the absorption, distribution, metabolism, excretion and toxicity properties of drug candidates. The drug-likeness model score for CID_4539974, CID_3642762 and CID_4268983 are 0.46, -0.64 and 0.58 respectively.



CID_4539974	CID_3642762	CID_42689838

Figure 20: Drug-likeness model score by Molsoft server.

The result of Molsoft and PreADMET analysis of best 3 molecules are shown in tabulated form below:

Table 7: Molsoft and PreADMET analysis of CID_4539974

	Molecular properties		ADME properties		Toxicity testing			
					Ames test		Carcinogenicity test	
	Test	Result	Test	Result	Test	Result	Test	Result
CID_4539974	Molecular fomula	$C_{12}H_{16}FN_3O_3S$	Human Intestinal absorbtion	92.52273	Ames TA100 (+S9)	-ve	Mouse	-ve
	Molecular weight	301.09	Caco-2 cell permeability	4.27	Ames TA100 (-S9)	-ve	Rat	-ve

No. of HBA	5	MDCK cell permeability	1.28573	Ames TA1535(+S9)	+ve
No. of HBD	3	Skin permeability	-4.85983	Ames TA1535 (-S9)	-ve
LogP	0.44	Plasma protein binding	46.13246	Ames TA98 (+S9)	-ve
Logs	-3.23	BBB penetration	0.14819	Ames TA98 (-S9)	-ve
TPSA	80.79 A^2				
Volume	259.58 A^3				

Table 8: Molsoft and PreADMET analysis of CID_3642762:

	Molecular properties		**ADME properties**		**Toxicity testing**			
					Ames test		**Carcinogenicity test**	
	Test	**Result**	**Test**	**Result**	**Test**	**Result**	**Test**	**Result**
CID_3642762	Molecular fomula	$C_{12}H_{15}N_5O_2S$	Human Intestinal absorbtion	97.49478	Ames TA100 (+S9)	-ve	Mouse	-ve
	Molecular weight	293.09	Caco-2 cell permeability	17.2673	Ames TA100 (-S9)	+ve	Rat	-ve
	No. of HBA	6	MDCK cell permeability	8.66081	Ames TA1535 (+S9)	+ve		
	No. of HBD	0	Skin permeability	-3.64644	Ames TA1535 (-S9)	-ve		
	LogP	1.59	Plasma protein binding	47.80738	Ames TA98 (+S9)	+ve		
	Logs	-2.82	BBB penetration	0.261164	Ames TA98 (-S9)	-ve		
	TPSA	71.71 A^2						
	Volume	255.20A^3						

Table 9: Molsoft and PreADMET analysis of CID_4268983:

	Molecular properties		**ADME properties**		**Toxicity testing**			
					Ames test		**Carcinogenicity test**	
	Test	**Result**	**Test**	**Result**	**Test**	**Result**	**Test**	**Result**

CID_4268983	Molecular fomula	$C_{13}H_{19}N_3O_4S$	Human Intestinal absorbtion	91.02455	Ames TA100 (+S9)	-ve	Mo-use	-ve
	Molecular weight	313.11	Caco-2 cell permeability	10.2067	Ames TA100 (-S9)	-ve	Rat	-ve
	No. of HBA	3	MDCK cell permeability	2.45932	Ames TA1535(+S9)	+ve		
	No. of HBD	3	Skin permeability	-4.5932	Ames TA1535 (-S9)	-ve		
	LogP	0.26	Plasma protein binding	28.07151	Ames TA98 (+S9)	-ve		
	Logs	-3.09	BBB penetration	0.109841	Ames TA98 (-S9)	-ve		
	TPSA	88.40 A^2						
	Volume	285.56 A^3						

The properties of the best 3 molecules are predicted from the above table. The predicted properties are listed below:

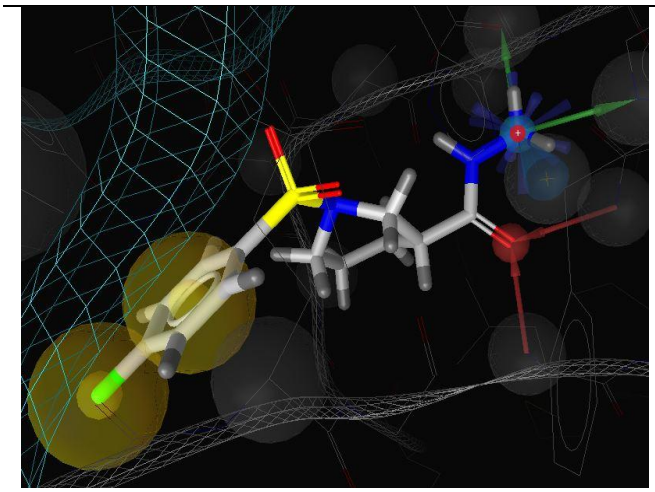
Table 10: Prediction of properties of molecules

Properties		**Molecule**		
		CID_4539974	**CID_3642762**	**CID_4268983**
Molecular	Deviation from lipinski rule	0	0	0
ADME	Human Intestinal absorbtion	Well	Well	Well
	Caco-2 cell permeability	Moderate	Moderate	Moderate
	MDCK cell permeability	Low	Low	Low
	Plasma protein binding	Weakly bound	Weakly bound	Weakly bound
	BBB penetration	Inactive	Inactive	Inactive
Toxicity	AMES	Negative	Positive	Negative
	Carcinogenicitty	Negative	Negative	Negative

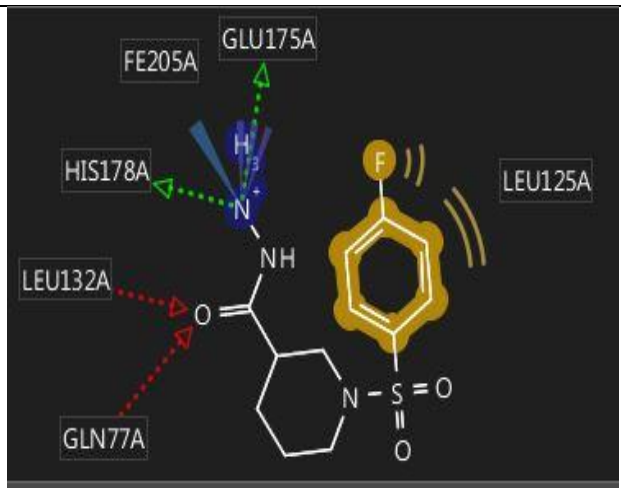
The above table demonstrate the properties of molecule as a potent drug candidate. None of molecule out of three violates the Lipinski rule of five. All molecules have well human intestinal absorption, moderate Caco-2 cell permeability, low MDCK cell permeability, weakly bound to plasma protein and inactive BBB penetration. None of these molecule show carcinogenicity and only one molecule CID_3642762 show positive AMES test. This result depicts that two molecules CID_4539974 and CID_4268983 have all desirable properties to be a drug. Combining and analyzing this result with docking and H-bond analysis result, 1-(4-fluorophenyl)sulfonylpiperidine-3-carbohydrazide (CID_4539974) showed to be a best potent PDI with having best docking result, inhibiting PDF of all selected microorganism, best H-bond analysis result, makes H-bond with active metal binding amino acid residue of PDF in all bacteria and containing all properties of drug molecule. Therefore, CID_4539974 was sorted as a best drug candidate out of 118 isomers of sulfonylpiperidine and selected for pharmacophore modelling and molecular dynamics simulation to study protein-ligand stability in biological system using GROMACS.

4.7 Pharmacophore modelling of CID_4539974-PDF Complex

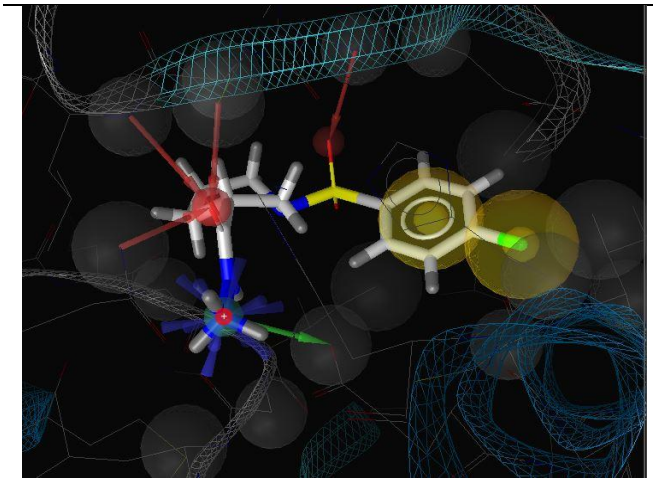
Pharmacophore modelling was done to identify the key features on the ligand which are interacting with points on the protein. It was done for two protein-ligand complex. One for PDF of *S.Mutans* (3L87) with which CID_4539974 gave highest binding energy of -8.75kcal/mol and other for PDF of *E.Coli* (1BSK) with which CID_4539974 gave lowest binding energy of -10.54 kcal/mol.



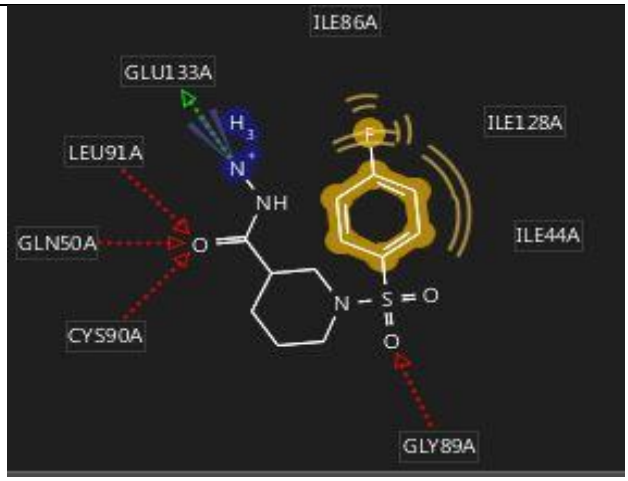
3L87-CID_4539974 complex 3-D view



3L87-CID_4539974 complex 2-D view



1BSK-CID4539974 complex 3-D view



1BSK-CID4539974 complex 2-D view

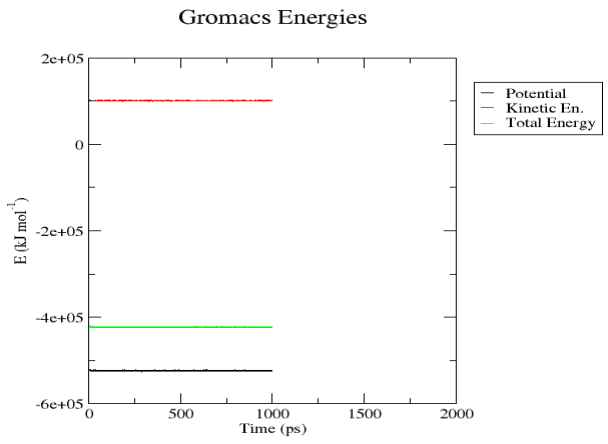
Figure 21: 3-D and 2-D pharmacophore model of 3L87-CID_4539974 complex and 1BSK-CID4539974 complex respectively.

The pharmacophore model of 3L87-CID_4539974 shows 2 H-Bond donors and 2 H-Bond acceptor. Leu132 is neighbour of active metal binding amino acid residue Cys131. While the pharmacophore model of 1BSK-CID_4539974 shows 1 H-Bond donors and 4 H-Bond acceptor. Cys90 is active metal binding amino acid residue.

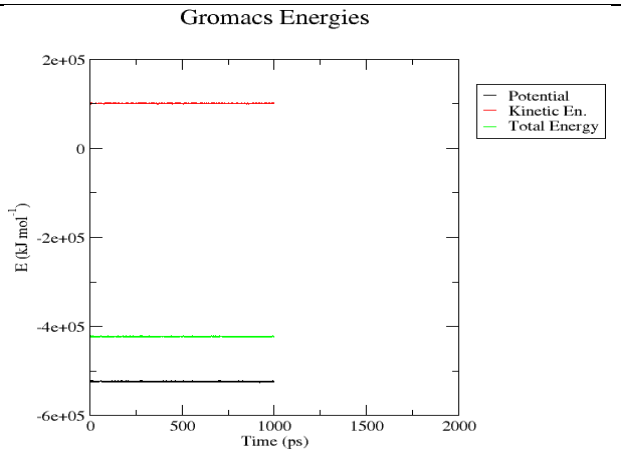
4.8 Molecular Dynamics Simulation of CID_4539974-PDF Complex:

Molecular dynamics simulation was run to study the stability of complex i.e. CID_4539974-PDF in biological system. This computational method calculated the time dependent behaviour of protein-ligand complex. Molecular Dynamics simulation has provided detailed information on the fluctuations and conformational changes of PDF. In order to validate our ligand first the PDF enzyme alone and then PDF-CID_4539974 complex was subjected to exhaustive a 1ns equilibrated Molecular Dynamics Simulation. System total energy, root mean square deviation (RMSD) and root mean square fluctuation (RMSF) file were generated for both protein alone and protein-ligand complex. MD simulation was run for two protein-ligand complex. One for PDF of *S.Mutans* (3L87) with which CID_4539974 gave lowest binding energy of -8.75kcal/mol and other for PDF of *E.Coli* (1BSK) with which CID_4539974 gave highest binding energy of -10.54 kcal/mol. We selected these two because on the basis of their result we can predict the simulation properties of remaining six protein-ligand complexes. All the results are shown in figure below:

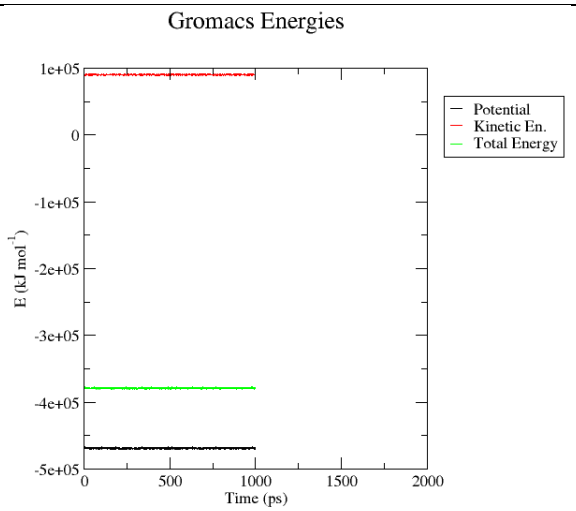
4.8.1 Gromacs energies result:



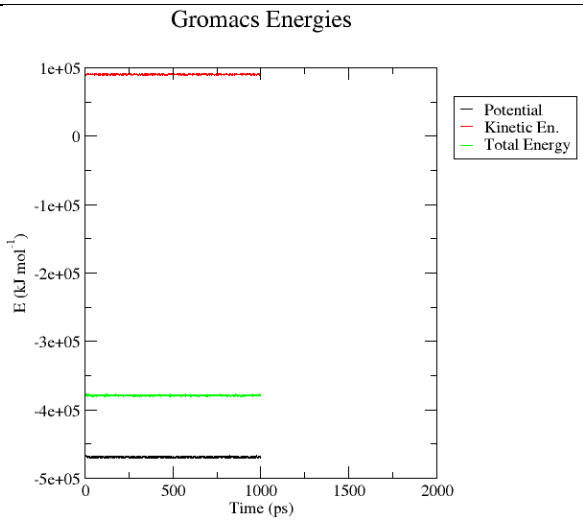
3L87 alone



3L87 – CID_4539974 complex



1BSK alone

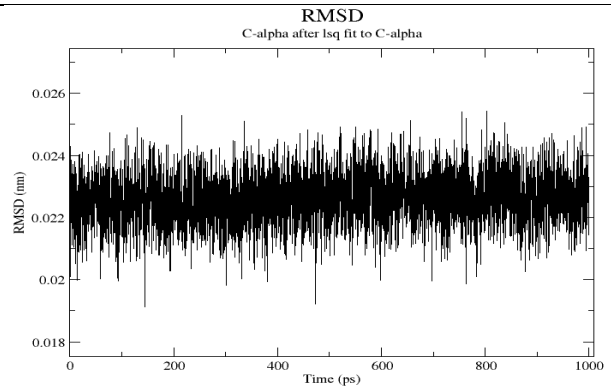


1BSK-CID4539974 complex

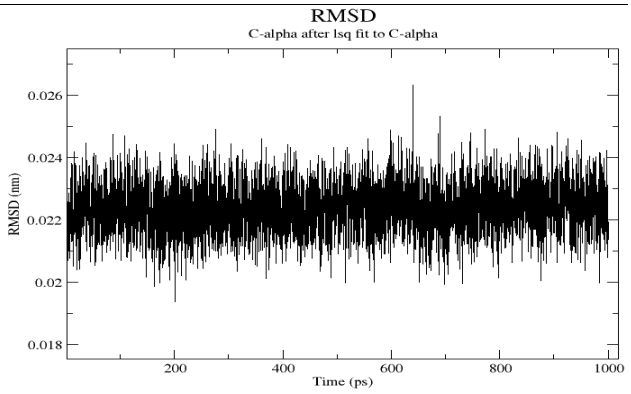
Figure 22: System energy of PDF (3L87 and 1BSK) without and with CID_4539974

The above Gromacs energies result shows that the total energy of the system remains almost constant after the binding of ligand with protein. Figure on the left side shows system energy of 3L87 and 1BSK when they were not bound with ligand while the figure on the right side represents the system energy of protein-ligand complex. The binding of ligand with protein does not cause significant change in system energy. For 3L87, the total energy of the system remains constant and fluctuates around -4.2×10^5 kJ/mol before and after binding of ligand, while for 1BSK, the total energy of system remains constant and fluctuates around -3.8×10^5 kJ/mol before and after binding of ligand.

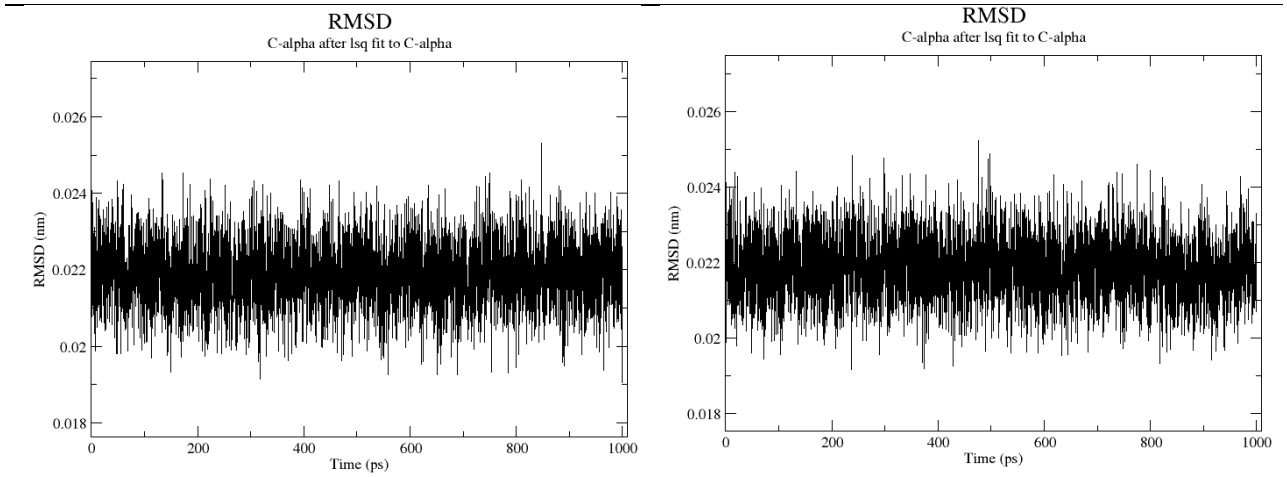
4.8.2 Gromacs Root Mean Square Deviation (RMSD) results:



3L87 alone

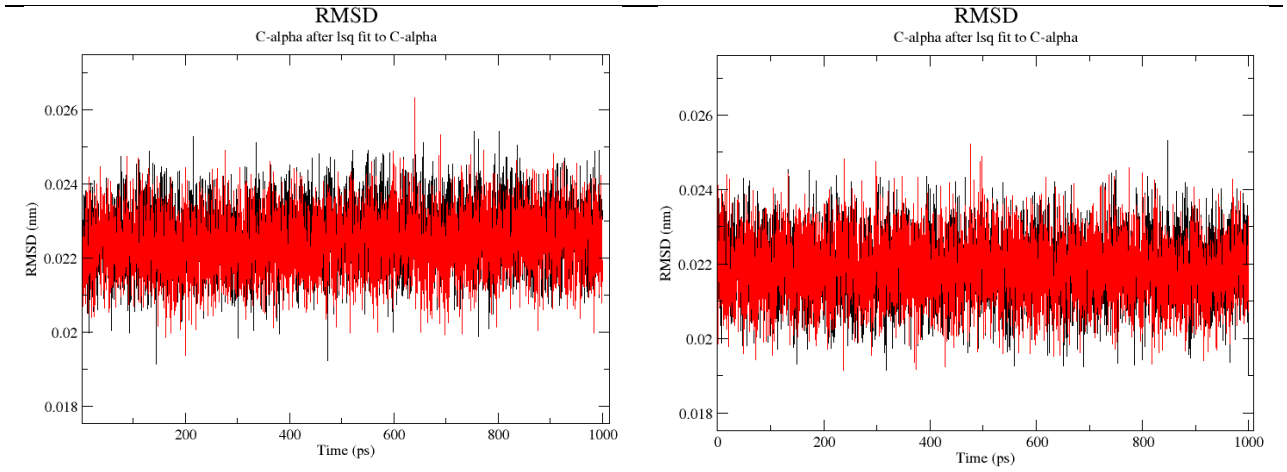


3L87 – CID_4539974



1BSK alone 1BSK-CID4539974 complex

Figure 23: RMSD with respect to simulation time of 1ns for 3L87 and 1BSK with and without CID_4539974



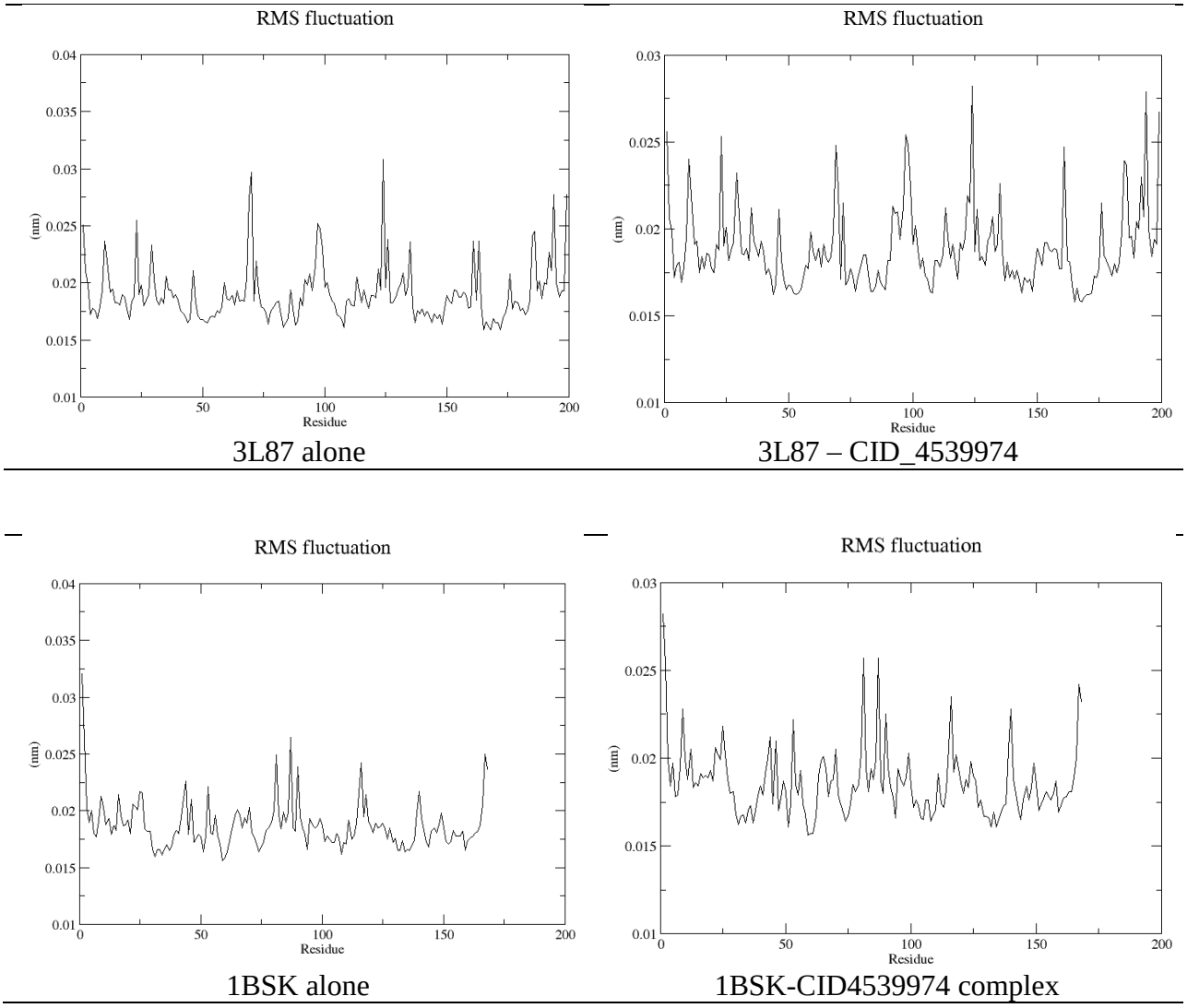
3L87 vs 3L87-CID_4539974 complex 1BSK vs 1BSK-CID_4539974 complex

Figure 24: Superposition of RMSD of 3L87 with 3L87-CID_4539974 complex and 1BSK with 1BSK-CID_4539974 complex.

Figure 23 shows the evolution of RMSD of Cα from the crystal structure during MD simulation. The protein atoms do not significantly deviate from the crystal structure and the Cα RMSD is stabilized to an average value of 0.022 ± .002 nm in both 3L87 and 1BSK before and after

binding of ligand. This result is well agreement with the result of figure y. In figure 24, superposition of RMSD of 3L87 with 3L87-CID_4539974 (left) and 1BSK with 1BSK-CID_4539974 complex (right) was done. Superposition shows there is negligible deviation in RMSD after binding of ligand with enzyme.

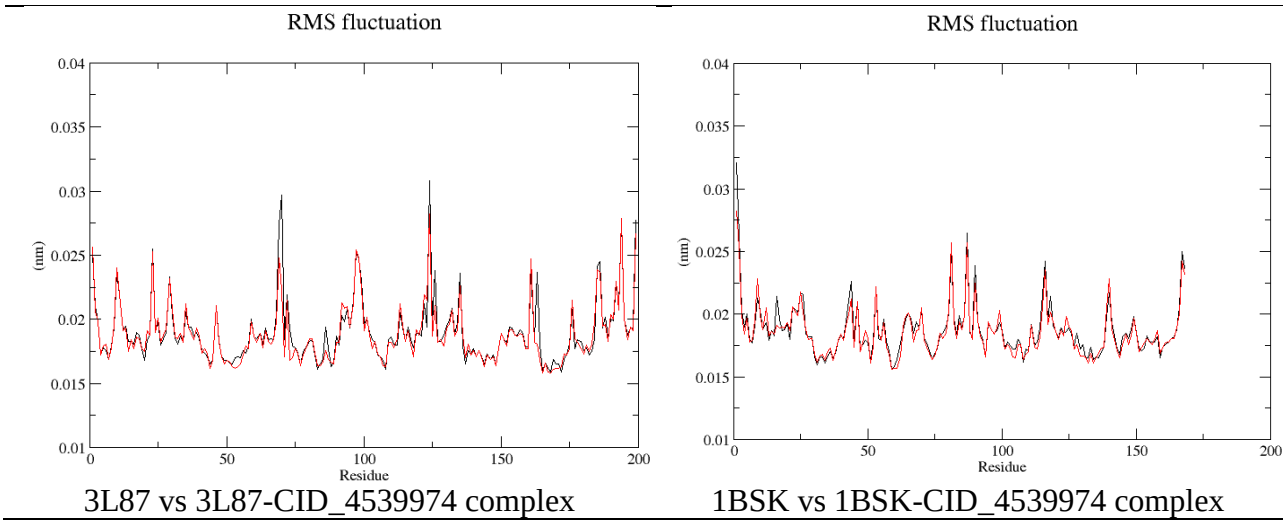
4.8.3 Gromacs Root Mean Square Fluctuation (RMSF) results:



3L87 alone | 3L87 – CID_4539974

1BSK alone | 1BSK-CID4539974 complex

Figure 25: RMSF with respect to residue during MD simulation for 3L87 and 1BSK with and without ligand.



3L87 vs 3L87-CID_4539974 complex 1BSK vs 1BSK-CID_4539974 complex

Figure 26: Superposition of RMSF of 3L87 with 3L87-CID_4539974 complex and 1BSK with 1BSK-CID_4539974 complex.

The flexibilities of the proteins were considered by the RMSF values obtained from conventional MD simulations which reflect the flexibility of each amino acid residue in the protein. Figure 25, shows the RMSF of the backbone of 3L87 and 1BSK with and without ligand as a function of residue number. The results indicate that the major fluctuations of residues occur in loops region connecting secondary structure elements. These residues are located in the exterior area of protein, and correspond to the regions of protein with large RMSD. In a characteristic RMSF pattern, a low RMSF value indicates the well-structured regions while the high values indicate the loosely structured loop regions or domains terminal. In the 3L87 and 1BSK major fluctuation occurs in region containing active metal binding amino acid residue i.e., Cys131 and Cys91 respectively.

In figure 26, superposition of RMSF graph of 3L87 with 3L87-CID_4539974 complex and 1BSK with 1BSK-CID_4539974 complex shows that there is decrease in RMSF value after binding of ligand with enzyme. This result demonstrates the stability of PDF-ligand complex.

CHAPTER 5

CONCLUSION

5. CONCLUSION

5.1 Summary

- In the current study, a novel inhibitor of bacterial PDF has been developed that can be used as a broad-spectrum antimicrobial since the active site of Peptide Deformylase that is composed of three sequence motifs, was confirmed to be conserved throughout the microorganism.
- Since PDF is essential for bacterial survival but apparently not important for animal cell, it provides an attractive target for broad-spectrum antibacterial.
- Since hydrazides have shown inhibitory properties for PDF, Sulfonylpiperidine, a hydrazine from the vast group of hydrazides, was considered for docking and the result confirmed its essential role in inhibiting PDF from high binding energy in all bacterial species tested.
- Among the 118 ligands tested CID_4539974 was found to be in the best 5 list in all bacteria's, CID_3642762 was found to be in the best 5 list in all except S. mutans (3L87) and S. aureus (1LQW), CID_4268983 was found to be in the best 5 list in all except P. aeruginosa (1N5N) and E.coli (1BSK).
- CID_4539974, CID_3642762 and CID_4268983 were selected for H-bond analysis, molecular properties calculation and ADME & toxicity prediction.
- The result obtained sorted 1-(4-fluorophenyl) sulfonylpiperidine-3-carbohydrazide (CID_4539974) to be best PDI out of three.
- CID_4539974 made H-bond with active metal binding amino acid residue of all bacterial PDF, didn't deviate from Lipinski rule, and exhibited best inhibition affinity for PDF, good drug-likeness model score, optimal intestinal absorption, moderate Caco-2 cell permeability, week bound to plasma protein binding, inactive BBB penetration, negative AMES and carcinogenic property.
- Finally, MD simulation checked the stability of protein-ligand complex under biological conditions and was run for highest and lowest binding energy protein-ligand complex of CID_4539974. Kinetic energy, RMSD and RMSF graph confirmed the stability of PDF-CID_4539974 complex in biological system.

5.2 Conclusion

Lower binding energies and inhibition constant of CID_4539974 than actinonin concluded that it has better inhibitory properties than actinonin. The molecular, ADME and toxicity of CID_4539974 established it has drug like properties and could be a potent PDI (Peptide Deformylase Inhibitor). MD simulation concluded that PDF-CID_4539974 complex is stable under biological condition. All these outcomes interpret that CID_4539974 have all the properties to become a broad spectrum antibacterial by targeting bacterial PDF efficiently.

5.3 Future perspectives

Rising cost of drug development, increased bacterial resistance, emerging new diseases and lack of enough therapeutic agents have been a major concern in recent years. Therefore we need to alternative methods to get solutions on safety and efficacy faster and at lower cost. One such method is *insilico* drug designing. It plays significant role in all stages of drug development from preclinical stage to late clinical stage. In this modern era of Computer Aided Drug Design (CADD), a novel work is first started by virtual screening or *insilico* designing, and then goes for invitro and invivo analysis. Development of new drug is a sequential process start with *insitu* study, *insilco* drug design, invitro analysis and finally invivo testing. The availability of bioinformatics tools and software helped to identify probable antimicrobial drug to dock with protein targets.

All the above findings are generated with the help of most reliable tools and software of computational biology. The ligands discussed in the current study need QSAR study before qualifying for in vitro analysis followed by clinical trial. Moreover, since the microorganism are fast gaining resistance to the existing drugs, the drug identified in the study,1-(4-fluorophenyl) sulfonylpiperidine-3-carbohydrazide (CID_4539974) can provide a better option for designing new drugs by modifying its chemical structure.

REFERENCES

1. Cohen, M.L., *Changing patterns of infectious disease.* Nature, 2000. **406**(6797): p. 762-767.
2. Yoneyama, H. and R. Katsumata, *Antibiotic resistance in bacteria and its future for novel antibiotic development.* Bioscience, biotechnology, and biochemistry, 2006. **70**(5): p. 1060-1075.
3. Fauci, A.S., *Infectious diseases: considerations for the 21st century.* Clinical Infectious Diseases, 2001. **32**(5): p. 675-685.
4. Nathan, C., *Antibiotics at the crossroads.* Nature, 2004. **431**(7011): p. 899-902.
5. Levy, S.B., *Antibiotic resistance: consequences of inaction.* Clinical Infectious Diseases, 2001. **33**(Supplement 3): p. S124-S129.
6. Olarte, J., *Antibiotic resistance in Mexico.* APUA Newsletter, 1983. **1**: p. 3ff.
7. Watanabe, T., *Infective heredity of multiple drug resistance in bacteria.* Bacteriological Reviews, 1963. **27**(1): p. 87.
8. Gross, M., et al., *Oral anti-pneumococcal activity and pharmacokinetic profiling of a novel peptide deformylase inhibitor.* Journal of Antimicrobial Chemotherapy, 2004. **53**(3): p. 487-493.
9. Hackbarth, C.J., et al., *N-alkyl urea hydroxamic acids as a new class of peptide deformylase inhibitors with antibacterial activity.* Antimicrobial agents and chemotherapy, 2002. **46**(9): p. 2752-2764.
10. Sakharkar, K.R., M.K. Sakharkar, and V.T. Chow, *A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa.* In silico biology, 2004. **4**(3): p. 355-360.
11. Luty, B.A., et al., *A molecular mechanics/grid method for evaluation of ligand–receptor interactions.* Journal of Computational Chemistry, 1995. **16**(4): p. 454-464.
12. Nguyen, K.T., *Mechanism, function, and inhibition of peptide deformylase*. 2005, Ohio State University.
13. Chan, P.F., et al., *Characterization of a novel fucose-regulated promoter (PfcsK) suitable for gene essentiality and antibacterial mode-of-action studies in Streptococcus pneumoniae.* Journal of bacteriology, 2003. **185**(6): p. 2051-2058.
14. Mazel, D., S. Pochet, and P. Marliere, *Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation.* The EMBO journal, 1994. **13**(4): p. 914.
15. Guay, D.R., *Drug forecast–the peptide deformylase inhibitors as antibacterial agents.* Therapeutics and clinical risk management, 2007. **3**(4): p. 513.
16. Fry, K.T. and M.R. Lamborg, *Amidohydrolase activity of< i> Escherichia coli</i> extracts with formylated amino acids and dipeptides as substrates.* Journal of molecular biology, 1967. **28**(3): p. 423-433.
17. Adams, J.M., *On the release of the formyl group from nascent protein.* Journal of molecular biology, 1968. **33**(3): p. 571-589.
18. Ben-Bassat, A., et al., *Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure.* Journal of bacteriology, 1987. **169**(2): p. 751-757.
19. Meinnel, T. and S. Blanquet, *Evidence that peptide deformylase and methionyl-tRNA (fMet) formyltransferase are encoded within the same operon in Escherichia coli*. Journal of bacteriology, 1993. **175**(23): p. 7737-7740.

20. Giglione, C., A. Boularot, and T. Meinnel, *Protein N-terminal methionine excision.* Cellular and Molecular Life Sciences CMLS, 2004. **61**(12): p. 1455-1474.
21. Chen, D. and Z. Yuan, *Therapeutic potential of peptide deformylase inhibitors.* 2005.
22. Meinnel, T., S. Blanquet, and F. Dardel, *A new subclass of the zinc metalloproteases superfamily revealed by the solution structure of peptide deformylase.* Journal of molecular biology, 1996. **262**(3): p. 375-386.
23. Becker, A., et al., *Iron center, substrate recognition and mechanism of peptide deformylase.* Nature Structural & Molecular Biology, 1998. **5**(12): p. 1053-1058.
24. Rajagopalan, P.R., S. Grimme, and D. Pei, *Characterization of cobalt (II)-substituted peptide deformylase: function of the metal ion and the catalytic residue Glu-133.* Biochemistry, 2000. **39**(4): p. 779-790.
25. Yoon, H.J., et al., *Crystal structure of peptide deformylase from Staphylococcus aureus in complex with actinonin, a naturally occurring antibacterial agent.* Proteins: Structure, Function, and Bioinformatics, 2004. **57**(3): p. 639-642.
26. Dardel, F., et al., *Solution structure of nickel-peptide deformylase.* Journal of molecular biology, 1998. **280**(3): p. 501-513.
27. Meinnel, T., *Peptide deformylase of eukaryotic protists: a target for new antiparasitic agents?* Parasitology today (Personal ed.), 2000. **16**(4): p. 165.
28. Giglione, C., M. Pierre, and T. Meinnel, *Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents.* Molecular microbiology, 2000. **36**(6): p. 1197-1205.
29. Boularot, A., et al., *Discovery and refinement of a new structural class of potent peptide deformylase inhibitors.* Journal of medicinal chemistry, 2007. **50**(1): p. 10-20.
30. Baldwin, E.T., et al., *Crystal structure of type II peptide deformylase from Staphylococcus aureus.* Journal of Biological Chemistry, 2002. **277**(34): p. 31163-31171.
31. Bracchi-Ricard, V., et al., *Characterization of an Eukaryotic Peptide Deformylase from< i> Plasmodium falciparum</i>.* Archives of biochemistry and biophysics, 2001. **396**(2): p. 162-170.
32. Park, J.K., et al., *Crystallization and preliminary X-ray crystallographic analysis of peptide deformylase (PDF) from Bacillus cereus in ligand-free and actinonin-bound forms.* Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 2004. **61**(1): p. 150-152.
33. Guilloteau, J.-P., et al., *The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: a platform for the structure-based design of antibacterial agents.* Journal of molecular biology, 2002. **320**(5): p. 951-962.
34. Zhou, Z., et al., *Unique Structural Characteristics of Peptide Deformylase from Pathogenic Bacterium< i> Leptospira interrogans</i>.* Journal of molecular biology, 2004. **339**(1): p. 207-215.
35. Mazel, D., et al., *A survey of polypeptide deformylase function throughout the eubacterial lineage.* Journal of molecular biology, 1997. **266**(5): p. 939-949.
36. Apfel, C., et al., *Hydroxamic acid derivatives as potent peptide deformylase inhibitors and antibacterial agents.* Journal of medicinal chemistry, 2000. **43**(12): p. 2324-2331.
37. Chen, D.Z., et al., *Actinonin, a naturally occurring antibacterial agent, is a potent deformylase inhibitor.* Biochemistry, 2000. **39**(6): p. 1256-1262.

38. Broughton, B.J., et al., *Studies concerning the antibiotic actinonin. Part VIII. Structure–activity relationships in the actinonin series.* J. Chem. Soc., Perkin Trans. 1, 1975(9): p. 857-860.
39. Clements, J.M., et al., *Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor.* Antimicrobial agents and chemotherapy, 2001. **45**(2): p. 563-570.
40. Gabr, M., S. Madathil, and Q. Hussain, *Molecular Docking & Analysis of Peptide Deformylase (PDF) with Hydrazides: Molecular Modeling Study of New Anti-Leptospiral Drugs.* AYRCOBE, 2012.
41. Antre, R.V., Kore, Pranita P., *Computer-Aided Drug Design: An Innovative Tool for Modeling.* Open Journal of Medicinal Chemistry, 2012, 2, 139-148, 2012.
42. Bharath, E., S. Manjula, and A. Vijaychand, *In silico drug design-tool for overcoming the innovation deficit in the drug discovery process.* Chemistry, 2011. **18**(23.3): p. 1.0.
43. White, S., *Pharma 2020: The Vision–Which path will you take?* SA Pharmaceutical Journal, 2007. **74**(9): p. 40-41.
44. Tang, Y., et al., *New technologies in computer-aided drug design: toward target identification and new chemical entity discovery.* Drug discovery today: technologies, 2006. **3**(3): p. 307-313.
45. Andricopulo, A.D., R.V. Guido, and G. Oliva, *Virtual screening and its integration with modern drug design technologies.* Current medicinal chemistry, 2008. **15**(1): p. 37-46.
46. Oprea, T.I. and H. Matter, *Integrating virtual screening in lead discovery.* Current opinion in chemical biology, 2004. **8**(4): p. 349-358.
47. Evers, A. and T. Klabunde, *Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor.* Journal of medicinal chemistry, 2005. **48**(4): p. 1088-1097.
48. Zhang, Q. and I. Muegge, *Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring.* Journal of medicinal chemistry, 2006. **49**(5): p. 1536-1548.
49. Lengauer, T., et al., *Novel technologies for virtual screening.* Drug discovery today, 2004. **9**(1): p. 27-34.
50. Johnson, M.A. and G.M. Maggiora, *Concepts and applications of molecular similarity*. Vol. 8. 1990: Wiley New York.
51. Willett, P., *Similarity-based virtual screening using 2D fingerprints.* Drug discovery today, 2006. **11**(23): p. 1046-1053.
52. Shoichet, B.K., *Virtual screening of chemical libraries.* Nature, 2004. **432**(7019): p. 862-865.
53. Klebe, G., *Virtual ligand screening: strategies, perspectives and limitations.* Drug discovery today, 2006. **11**(13): p. 580-594.
54. Gorse, D. and R. Lahana, *Functional diversity of compound libraries.* Current opinion in chemical biology, 2000. **4**(3): p. 287-294.
55. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
56. Goodsell, D.S. and A.J. Olson, *Automated docking of substrates to proteins by simulated annealing.* Proteins: Structure, Function, and Bioinformatics, 1990. **8**(3): p. 195-202.

57. Huey, R., et al., *A semiempirical free energy force field with charge-based desolvation.* Journal of Computational Chemistry, 2007. **28**(6): p. 1145-1152.
58. Lutz, M., D. Ascher, and F. Willison, *Learning python*. Vol. 2. 1999: O'Reilly.
59. Sanner, M.F., et al. *Integrating computation and visualization for biomolecular analysis: an example using Python and AVS*. in *Pac Symp Biocomput*. 1999.
60. Allen, M. and D. Tildesley, *Oxford University Press; New York: 1987.* Computer Simulation of Liquids.
61. Haile, J.M., *Molecular dynamics simulation: elementary methods*. 1992: John Wiley & Sons, Inc.
62. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural & Molecular Biology, 2002. **9**(9): p. 646-652.
63. Norberg, J. and L. Nilsson, *Molecular dynamics applied to nucleic acids.* Accounts of chemical research, 2002. **35**(6): p. 465-472.
64. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.* Accounts of chemical research, 2000. **33**(12): p. 889-897.
65. Straatsma, T. and J. McCammon, *Computational alchemy.* Annual Review of Physical Chemistry, 1992. **43**(1): p. 407-435.
66. Simonson, T., G. Archontis, and M. Karplus, *Free energy simulations come of age: protein-ligand recognition.* Accounts of chemical research, 2002. **35**(6): p. 430-437.
67. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chemical reviews, 2006. **106**(5): p. 1589.
68. Yoo, J.-S., et al., *Macrolactin N, a new peptide deformylase inhibitor produced by< i> Bacillus subtilis</i>*. Bioorganic & medicinal chemistry letters, 2006. **16**(18): p. 4889-4892.
69. Wolber, G. and T. Langer, *LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters.* Journal of chemical information and modeling, 2005. **45**(1): p. 160-169.
70. Kapupara, P., et al., *Journal of Chemical and Pharmaceutical Research.* J. Chem, 2010. **2**(3): p. 287-294.
71. Rout, P.R., R. Satpathy, and G.R. Satpathy, *Journal of Chemical and Pharmaceutical Research, 2012, 4 (2): 1200-1206.* Journal of Chemical and Pharmaceutical Research, 2012. **4**(2): p. 1200-1206.